



Universidade do Porto
Faculdade de Medicina
2012

**FILIPA ALEXANDRA
DE JESUS TAVARES**

**COMPARAÇÃO DE METODOLOGIAS NO ESTUDO DE SNPs
MITOCONDRIAIS COM INTERESSE FORENSE COM BASE
NUMA AMOSTRA DA POPULAÇÃO DE TIMOR-LESTE**

Dissertação de Mestrado em Ciências Forenses



Universidade do Porto
Faculdade de Medicina
2012

**FILIPA ALEXANDRA
DE JESUS TAVARES**

**COMPARAÇÃO DE METODOLOGIAS NO ESTUDO DE SNPs
MITOCONDRIAIS COM INTERESSE FORENSE COM BASE
NUMA AMOSTRA DA POPULAÇÃO DE TIMOR-LESTE**

Dissertação de candidatura ao grau de Mestre em Ciências Forenses
submetida à Faculdade de Medicina da Universidade do Porto.

Orientador: Doutor Luís Souto de Miranda

Categoria: Professor Auxiliar Convidado

Afiliação: Departamento de Biologia, Universidade de Aveiro

Co-orientadora: Doutora Maria de Fátima Terra Pinheiro

Categoria: Professor afiliada

Afiliação: ICBAS

O presente trabalho, realizado no Departamento de Biologia da Universidade de Aveiro no âmbito do projecto da FCT: PTDC/CS-ANT/108558/2008 TIMORPOP GENÉTICA DE TIMOR-LESTE, contou com o apoio do Programa Operacional Temático, Factores de Competitividade (COMPETE), Quadro Comunitário de Apoio III e FEDER

agradecimentos

Gostaria de agradecer a todos que me apoiaram e ajudaram a realizar este trabalho. Em particular:

Ao meu orientador, Doutor Luís Souto, pela disponibilidade em deixar-me realizar a dissertação no laboratório que dirige no Departamento de Biologia da Universidade de Aveiro e pela paciência e apoio. O meu obrigada.

À minha co-orientadora, Doutora Maria de Fátima Pinheiro, pelo incentivo e tempo partilhado.

Ao Departamento de Biologia da Universidade de Aveiro, agradeço a autorização para a realização prática desta dissertação e meios concedidos.

À delegação do Centro do Instituto Nacional de Medicina Legal, agradeço a autorização para utilizar os equipamentos do Serviço de Genética e Biologia Forense.

A todas as pessoas do Serviço de Genética e Biologia Forense da Delegação de Coimbra do Instituto Nacional de Medicina Legal, pela simpatia e disponibilidade, com que acolheram e ajudaram neste trabalho.

À minha amiga Helena Moreira pela paciência e por toda a ajuda nos momentos de mais necessidade e por estar sempre ao meu lado para tudo.

Aos meus colegas de curso e de laboratório.

Agradeço o apoio concedido pela Fundação para a Ciência e Tecnologia para a realização prática deste trabalho, através do projecto PTDC/CS-ANT/108558/2008 – “TIMORPOP GENÉTICA DE TIMOR-LESTE”.

Ao Samuel, obrigada por todo o carinho, apoio e toda a ajuda dada ao longo deste processo.

Finalmente, o meu maior e mais sincero obrigada aos meus pais e irmãos pelo apoio incondicional.

Obrigada a todos!

palavras-chave

PCR em Tempo Real; Minissequenciação; SNPs; Região codificante do DNA mitocondrial; População de Timor-Leste; Genética Forense

resumo

O DNA Mitocondrial humano (DNAMt) é um marcador especialmente útil para detectar a ancestralidade biogeográfica materna e é largamente utilizado em estudos forenses e antropológicos devido às suas características particulares.

Na Genética Forense, existe uma tendência crescente em usar o DNAMt quando estão disponíveis quantidades limitadas de DNA. A análise da região controladora do DNAMt tem a desvantagem de conter um poder limitado de discriminação, devido à falta de recombinação que resulta da herança uniparental materna. De forma a aumentar o poder de discriminação do DNAMt, nos últimos anos têm vindo a ser desenvolvidas novas metodologias de detecção de variações (SNPs) na região codificante do DNAMt, tais como a minissequenciação e PCR em Tempo Real, através da análise pós-PCR *High-Resolution Melting* (HRM).

Neste trabalho foi estudada uma amostra populacional de 38 indivíduos naturais de Timor-Leste, recorrendo a marcadores bi-alélicos localizados na região codificante do DNA mitocondrial, tendo como objectivos principais: (i) comparar duas metodologias de genotipagem de SNPs (1. PCR em Tempo Real, através da análise HRM; e 2. Minissequenciação, com a utilização do kit SNaPshot®) e discutir suas vantagens e limitações com interesse forense e (ii) determinar os haplogrupos presentes na amostragem populacional, de acordo com polimorfismos presentes na região HVS I e HVS II e região codificante do DNA mitocondrial.

Da comparação das duas metodologias de genotipagem de SNPs utilizadas neste estudo, verificou-se que a minissequenciação se revela uma metodologia mais precisa, robusta, fiável e com possibilidade de análise de vários *loci* em simultâneo, podendo esta metodologia ser utilizada nos laboratórios de Genética Forense. Contudo, a análise HRM também poderá ser utilizada como um método de pré-selecção das amostras forenses, antes da genotipagem por minissequenciação.

A determinação dos haplogrupos presentes nas amostras permitiu a obtenção de resultados que podem constituir uma aproximação a um futuro estudo populacional.

keywords

Real-Time PCR; Minisequencing; SNPs; Coding region of mitochondrial DNA; Population of East Timor; Forensic Genetics

abstract

Human Mitochondrial DNA (mtDNA) is an especially useful marker to detect maternal biogeographic ancestry and is widely used in forensic and anthropological studies due to its particular characteristics.

In Forensic Genetics, there is a rising tendency to use mtDNA when limited quantities of DNA are available. The analysis of the control region of mtDNA has the disadvantage of having a limited power of discrimination due to the lack of recombination which results from maternal uniparental inheritance. In order to increase the discrimination power of mtDNA, new methodologies used in the detection of variations (SNPs) in the coding region of mtDNA have been developed in recent years. Such methodologies include minisequencing and Real Time PCR, through the analysis of post-PCR High Resolution Melting (HRM).

In this work, a population sample of 38 individuals from East Timor was studied, using biallelic markers localized in the coding region of Mitochondrial DNA. The work had the following objectives: (i) Compare two SNP genotyping methodologies (1. Real-time PCR, through the analysis of HRM; and 2. minisequencing, using the SnaPshot® kit) and discuss their advantages and limitations in a forensic context; and (ii) determine the haplogroups present in the population sample, according to polymorphisms present in the HVS I and HVS II region and the coding region of Mitochondrial DNA.

Through the comparison of the two genotyping methodologies used in this study, it was verified that minisequencing is a more precise, robust, reliable methodology, capable of analyzing various loci simultaneously, such that it may be used in Forensic Genetics Laboratories. However, HRM analysis may also be used as a method in the pre-selection of forensic samples, before the genotyping through minisequencing process.

The determination of haplogroups present in the samples generated results that can constitute an approximation to a future population study.

ÍNDICE

1. Introdução.....	1
1.1. Timor-Leste	1
1.1.1. Contexto Histórico	2
1.2. Aplicação do estudo do DNA em Genética Forense	3
1.2.1. Marcadores Uniparentais	4
1.2.2. DNA Mitocondrial.....	5
1.2.2.1. <i>Single Nucleotide Polymorphism</i> (SNP)	6
1.2.2.2. Filogeografia mitocondrial.....	8
1.3. Metodologias de Genotipagem de SNPs.....	10
1.3.1. Minissequenciação – Extensão de <i>Primer</i>	10
1.3.2. PCR em Tempo Real	12
Análise HRM (<i>High-Resolution Melting Analysis</i>)	14
1.3.3. Comparação das metodologias: Vantagens e Desvantagens	16
2. Objetivos.....	19
3. Material e Métodos.....	21
3.1. Amostragem populacional.....	21
3.2. Extração de DNA.....	21
3.3. Genotipagem de SNPs.....	22
3.3.1. Minissequenciação	22
3.3.2. PCR em Tempo Real – Análise HRM	28
4. Resultados e Discussão	31
4.1. Minissequenciação	31
4.1.1. SNaPshot™ <i>Multiplex</i> kit – Seleção do polímero	31
4.1.2. Resultados.....	32
4.2. PCR em Tempo Real: Análise HRM	35
4.2.1. Análise comparativa com os três kits de amplificação.....	35
4.2.2. Resultados.....	36
4.3. Análise comparativa das metodologias.....	40
4.4. Distribuição de haplogrupos	41
5. Conclusões.....	43
6. Referências Bibliográficas.....	45

7. Anexos	51
Anexo I – Extração de DNA pelo método adaptado de <i>Chelex</i> ®100	51
Anexo II – Condições de amplificação do <i>Multiplex</i> PCR Master Mix (Qiagen)	52
Anexo III – Condições de amplificação do ABI Prism® SNaPshot™ <i>Multiplex</i> Kit (Applied Biosystems)	53
Anexo IV – Condições de amplificação dos kits utilizados para a análise HRM	54
Anexo V – Proposta de artigo	55

ÍNDICE DE FIGURAS

Figura 1 – Localização Geográfica de Timor-Leste.	1
Figura 2 - Diversidade linguística em Timor-Leste.....	2
Figura 3 – Representação esquemática do modo de transmissão do DNA.	4
Figura 4 - Representação esquemática do genoma mitocondrial humano.	5
Figura 5 – Variações existentes no DNA.....	6
Figura 6 – <i>Single Nucleotide Polymorphism</i> (SNP).....	7
Figura 7 – Representação esquemática da filogenia do DNA mitocondrial.	9
Figura 8 – Esquema representativo do método de minissequenciação.	11
Figura 9 – Gráfico de amplificação da PCR em Tempo Real.....	12
Figura 10 – Esquema ilustrativo de dois grandes métodos químicos de fluorescência.	13
Figura 11 – Figura ilustrativa de uma curva de <i>Melting</i>	14
Figura 12 – Esquema ilustrativo do modo de atuação dos corantes não saturados (A) e corantes saturados (B) durante a dissociação das cadeias de DNA.	15
Figura 13 – Árvore filogenética e nomenclatura dos haplogrupos do DNA mitocondrial definidos pelos 32 SNPs estudados.	23
Figura 14 – Eletroferogramas dos marcadores 3547/1736 do <i>Multiplex</i> I e 4820/4883 do <i>Multiplex</i> II com o polímero POP-4.....	32
Figura 15 – Representação gráfica da curva de amplificação das amostras teste nos três kits.....	35
Figura 16 – Exemplo das curvas de <i>melting</i> normalizadas e sua respetiva tabela de identificação para o marcador 4491, obtidas através da análise HRM.....	36
Figura 17 – Exemplo de resultados obtidos para o marcador 2092 com $T_a=56^{\circ}\text{C}$	38
Figura 18 – Resultados obtidos após o gradiente de temperatura para o marcador 2092.....	39
Figura 19 – Curva de <i>melting</i> normalizada, gráfico de diferenças obtido da curva de <i>melting</i> normalizada e tabela de identificação para as amostras 28, 49 e 62.....	40
Figura 20 – Comparação da sequência de referência com os <i>primers</i> utilizadas para detetar o polimorfismo 8414.....	41

ÍNDICE DE TABELAS

Tabela 1 – Tabela comparativa das metodologias: Minissequenciação e Análise HRM.	16
Tabela 2 - Sequência dos <i>primers</i> utilizados para a amplificação de SNPs.	25
Tabela 3 - Sequência dos <i>primers</i> utilizados para a reação de minissequenciação.....	27
Tabela 4 – Sequência dos <i>primers</i> utilizados para a amplificação e detecção de SNPs na PCR em Tempo Real.....	29
Tabela 5 – Tabela com os resultados da Minissequenciação da região codificante e sequenciação das regiões HVS I e HVS II.....	33
Tabela 6 – Resultados obtidos na PCR em Tempo Real, através da análise HRM.	37
Tabela 7 – Tabela comparativa dos resultados de 3 marcadores no SNaPshot™ e na análise HRM.	40

LISTA DE ABREVIATURAS

A – Adenina

ATP – Adenosina TriFosfato

C – Citocina

CRS - Sequência referência de Cambridge

ddNTP – Dideoxynucleotídeos

D-loop – *Displacement loop*

DNA – Deoxyribonucleic acid - Ácido desoxirribonucleico (ADN)

DNAmt – Ácido desoxirribonucleico mitocondrial

G – Guanina

HLA – Antígenos dos leucócitos humanos

HRM – *High-Resolution Melting*

HVS – Segmentos hipervariáveis

NRV – Região não codificante do cromossoma Y (do inglês Non-recombining region of Y chromosome)

Pb – Pares de base

PCR – Reação em cadeia da polimerase (do inglês *Polymerase Chain Reaction*)

POP – *Performance Optimized Polymer*

rCRS – Sequência referência revista de Cambridge

RDTL – República Democrática de Timor-Leste

RFLP – *Restriction Fragment Length Polymorphism*

RFU – Unidade de fluorescência

SAP – *Shrimp Alkaline Phosphatase*

SBE – Extensão de base única (do inglês *Single Base extension*)

SNP – Polimorfismo de nucleótido único (do inglês Single Nucleotide Polymorphism)

STR – Repetição curta em *tandem* (do inglês Short Tandem Repeat)

T – Timina

Tm – Temperatura de *melting*

VNTRs – *Variable Number of Tandem Repeat*

1. INTRODUÇÃO

1.1. Timor-Leste

Timor é uma ilha localizada no Sudoeste Asiático (a cerca de 550 km a norte da Austrália) e que se encontra dividida politicamente em duas partes: Timor Oeste (ou *Nusa Tenggara Timur*), que pertence a uma província da Indonésia; e Timor-Leste, que ocupa a parte oriental da ilha de Timor (Figura 1).

Timor-Leste, também conhecido como República Democrática de Timor-Leste (RDTL), é a menor e a mais oriental das ilhas do arquipélago malaio. Este é constituído pela ilha de Ataúro (a norte da costa de Díli), o ilhéu de Jaco (no extremo leste) e o enclave de Oecussi-Ambeno, na costa norte da parte indonésia da Ilha de Timor.

O território de Timor-Leste tem uma área total de aproximadamente 15 mil km², com uma população de cerca de 1 milhão de habitantes, e tem por capital a cidade de Díli (SEKM, 2011).



Figura 1 – Localização Geográfica de Timor-Leste.
(Adaptado de Vidiani (2011))

A RDTL é o único país independente na Ásia de língua oficial portuguesa e encontra-se dividido em 13 distritos: Bobonaro, Liquiçá, Díli, Baucau, Manatuto e Lautém na costa norte; Cova-Lima, Ainaro, Manufahi e Viqueque na costa sul; Ermera e Aileu, situados no interior montanhoso; e Oecussi-Ambeno, enclave no território indonésio.

Devido ao seu contexto histórico e à sua localização geográfica, Timor-Leste apresenta uma elevada diversidade linguística (Figura 2) demonstrando pelo menos dezasseis línguas nativas (Hattori, Gomes, Ajo, & Belo, 2005; Hull, 2004).

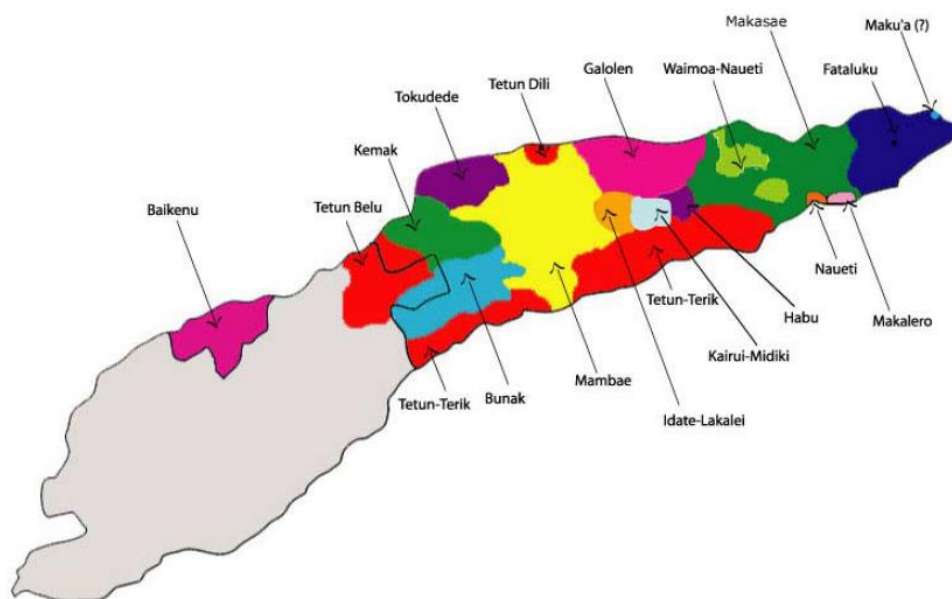


Figura 2 - Diversidade linguística em Timor-Leste.
(Adaptado de Fox and Soares (2003))

1.1.1. Contexto Histórico

A história da ocupação humana na ilha de Timor data desde há cerca de 30.000 – 35.000 anos, segundo investigações arqueológicas à gruta Lene Hara, na ponta Este de Timor perto de Tutuala (O'Connor, Spriggs, & Veth, 2002).

O primeiro povo a chegar a Timor, aproximadamente entre 40.000 e 20.000 anos a.C., foi o dos Vedo-Australóides, semelhantes aos Vedas do Ceilão. Seguiram-se os Melanésios por volta de 3.000 anos a. C., semelhantes aos que atualmente vivem em Papua Nova-Guiné e em algumas ilhas do Pacífico. Uma terceira vaga de imigrantes deu-se em 2.500 a.C. com os chamados ‘proto-malaiois’ – povos oriundos do Sul da China e da Indochina do Norte (O'Connor et al., 2002; UNDP, 2002).

A ilha de Timor atraiu comerciantes chineses e malaiois, devido à sua abundância em mel, cera e principalmente sândalo (madeira nobre utilizada na perfumaria e móveis de luxo, que cobria praticamente toda a ilha). Os portugueses, atraídos por esses recursos naturais, chegaram à costa de Timor onde agora se encontra o enclave Oecussi por volta de 1515 (UNDP, 2002).

Quando os primeiros Portugueses chegaram a Timor, encontraram a ilha dividida em dois reinos: Senobai, na parte oeste da ilha (correspondendo atualmente a Timor Indonésio) e Behale, na parte leste (correspondendo atualmente a Timor-Leste). Nessa altura os contactos iniciais eram muito diminutos, limitando-se a algumas trocas comerciais em localidades junto à costa. Só no século XVII, devido ao comércio de sândalo

(que despertava interesse económico), é que os Portugueses se estabeleceram em algumas zonas costeiras, fixando-se especialmente comerciantes e missionários (Branco, 1987).

Pouco tempo depois, a ilha foi palco da rivalidade luso-holandesa pelo controlo do comércio de especiarias no Sudeste Asiático, fazendo com que os Portugueses estabelecessem uma presença mais significativa na ilha. A disputa foi solucionada através de vários acordos, que culminaram com o Tratado das Fronteiras de 1916, terminando com as disputas territoriais e traçando a divisão da ilha entre Timor Ocidental Holandês e Timor Oriental Português.

Sendo um território estratégico entre a Austrália, Indonésia e Filipinas; Timor foi tomado durante a II Guerra Mundial pelos australianos que pretendiam organizar uma resistência no território. Em Fevereiro de 1942, o Japão usou a presença dos australianos como pretexto para entrar em Timor, onde permaneceu até Setembro de 1945, apesar da declaração de neutralidade de Portugal no conflito (UNDP, 2002).

Em 1945, a administração portuguesa foi restaurada no Timor Português, seguindo-se um período de quase três décadas em que não se manifestaram movimentos independentistas.

A 28 de Novembro de 1975, após uma breve guerra civil, a RDTL foi proclamada. Aproveitando o momento de tensão civil na antiga colónia portuguesa e incentivado pelas potências ocidentais, a nova nação foi ocupada pela Indonésia a 7 de Dezembro de 1975, que a ocupou durante 24 anos seguintes (Centeno & Novais, 2006).

Em 30 de Agosto de 1999, os timorenses votaram por esmagadora maioria pela independência, pondo fim a 24 anos de ocupação indonésia, na sequência de um referendo promovido pelas Nações Unidas. A 20 de Maio de 2002, Timor-Leste tornou-se a mais recente democracia do mundo e o primeiro estado-nação a ser constituído no século XXI (Centeno & Novais, 2006).

1.2. Aplicação do estudo do DNA em Genética Forense

A maioria da sequência de DNA humano (cerca de 99.7%) é a mesma entre indivíduos. Apenas uma porção do DNA difere entre indivíduos, tornando-os únicos. Através das regiões variáveis do genoma humano é possível a utilização dessa informação para a identificação humana (Butler, 2009).

Em 1985, o geneticista inglês Alec Jeffreys e seus colaboradores descreveram pela primeira vez um método de identificação individual que designaram por '*DNA fingerprinting*' ou tipagem do DNA, como é conhecido atualmente (Jeffreys, Wilson, & Thein, 1985). Este método foi um elemento chave para o desenvolvimento da Genética Forense (Carracedo, 2005).

Alec Jeffreys descobriu que certas regiões do DNA repetiam-se diversas vezes e que o número de repetições era variável entre indivíduos. Estas regiões ficaram conhecidas como minissatélites (também designado de VNTRs, do inglês *variable number of tandem repeats*). A técnica utilizada para estudar os VNTRs foi designada por *restriction fragment length polymorphism* (RFLP), que envolvia o uso de enzimas de restrição para detetar loci polimórficos (Butler, 2009).

Antes da introdução do '*DNA fingerprint*', os casos de Genética Forense (e.g., testes de paternidade, casos criminais e identificação individual) eram realizados utilizando marcadores serológicos clássicos, dos quais se destacam: os grupos sanguíneos, os antígenos dos leucócitos humanos (HLA) e as proteínas e enzimas polimórficas (cuja tipagem era efetuada usando metodologias imunológicas ou eletroforéticas).

A análise dos minissatélites foi um grande passo para a evolução da Genética Forense, mas apresentava várias limitações, tais como: exigir uma quantidade considerável de DNA (nanogramas a microgramas); não proporcionar resultados em amostras degradadas; a interpretação e valorização estatística dos resultados ser difícil e, ser uma técnica demorada e trabalhosa (Pinheiro, 2010).

Em 1985, foi apresentada uma nova técnica – a *polymerase chain reaction* (PCR) – que revolucionou a área da Biologia Molecular, incluindo a da Genética Forense (Mullis et al., 1986). Esta técnica, descrita pela primeira vez em 1985 por Kary Mullis e colaboradores, tornou possível a obtenção de um elevado número de cópias de uma sequência específica de DNA em poucas horas, obviando as limitações dos RFLPs.

1.2.1. Marcadores Uniparentais

Na transmissão de DNA de uma geração para a outra, pode ocorrer recombinação genética (*crossing-over*) entre cromossomas homólogos, com consequente troca de material genético, tornando cada pessoa única. No entanto, o mesmo não acontece em relação à região não codificante do cromossoma Y (NRY) e ao DNA mitocondrial que, devido à ausência de recombinação, são transmitidos em blocos, permanecendo praticamente inalterados (exceto eventos mutacionais) à medida que passam de geração para geração. A linhagem materna pode ser detetada pela informação de polimorfismos do DNA mitocondrial, enquanto que a linhagem paterna pode ser seguida pelos marcadores do cromossoma Y (Figura 3). (Butler, 2009; Human_Genome_Project_Information, 2010).

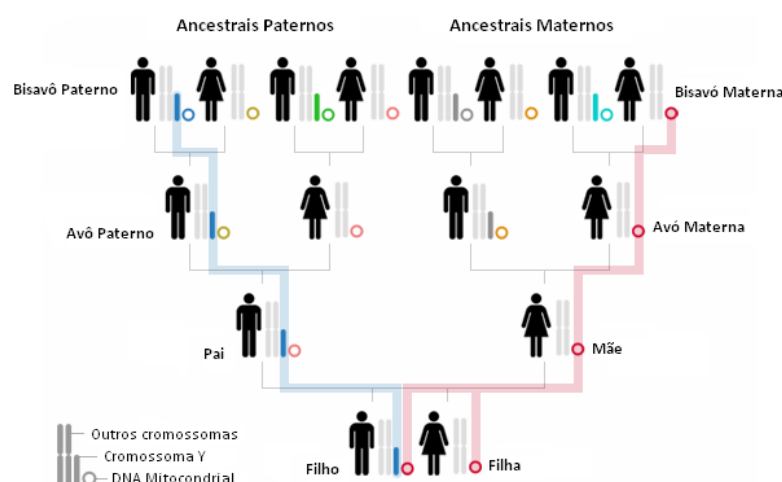


Figura 3 – Representação esquemática do modo de transmissão do DNA.
(Adaptado de National_Geographic (1996))

Os marcadores genéticos presentes no DNA mitocondrial e no cromossoma Y, designados por marcadores uniparentais ou de linhagem, contêm polimorfismos que

permitem inferir a ancestralidade através de várias gerações, ao contrário dos marcadores autossómicos que sofrem recombinação.

As características de transmissão dos marcadores uniparentais tornam-nos particularmente úteis em diversas áreas, tais como antropologia evolucionária, histórias de populações, genética médica e Genética Forense. Estes marcadores permitem definir linhagens masculinas e femininas (de informação genética), sendo que a análise dos respetivos padrões de distribuição populacional possibilita inferir sobre aspetos importantes da história demográfica das populações e discriminar os contributos, de ambos os sexos, no que concerne a genética de populações (Kayser, 2007).

Relativamente à Genética Forense, a análise de marcadores específicos do cromossoma Y e do DNA mitocondrial é especialmente útil em criminalística biológica.

1.2.2. DNA Mitocondrial

O DNA mitocondrial (DNAm_t) representa uma pequena porção do genoma humano (1-2 %) encontrando-se em organelos celulares específicos – as mitocôndrias – cuja principal função prende-se com a produção de energia na forma de ATP (Adenosina TriFosfato), necessária para o bom funcionamento da célula.

O genoma mitocondrial humano (Figura 4) consiste numa molécula circular de dupla cadeia [uma cadeia leve, L (*Light*), rica em Citocinas (C) e Timinas (T) e uma cadeia pesada, H (*Heavy*), rica em Guaninas (G) e Adeninas (A)], com um comprimento aproximado de 16569 pares de bases (pb), podendo o valor final de nucleótidos no genoma mitocondrial variar devido a mutações de um pequeno número de nucleótidos, que podem ser inserções ou deleções (Anderson et al., 1981; Andrews et al., 1999).

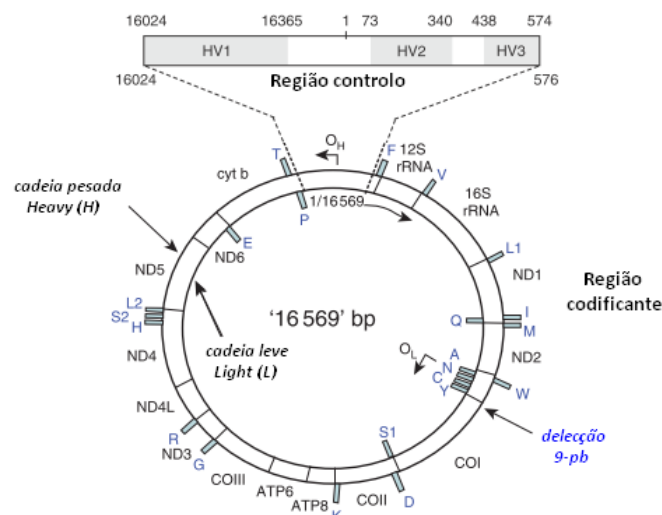


Figura 4 - Representação esquemática do genoma mitocondrial humano.
(Adaptado de Butler (2009))

O genoma mitocondrial é composto por uma região codificante, com cerca de 15500 pb, e uma região não codificante (denominada região controlo ou D-loop – *displacement loop*), com aproximadamente 1100 pb.

Na região controlo, situam-se os segmentos com maior variabilidade do genoma mitocondrial, que são denominados de segmentos hipervariáveis (HVS): o segmento hipervariável I (HVS I) que corresponde à região entre as posições 16024 e 16365; o segmento hipervariável II (HVS II), com 268 pb, que está localizado entre as posições 73 e 340, e o segmento hipervariável III (HVS III) que corresponde à região entre as posições 483 e 574. O terceiro segmento (HVS III) é o menos polimórfico dos segmentos, sendo utilizado em alguns casos, quando é necessária informação adicional (Butler, 2009; VanOven & Kayser, 2009)

A taxa de mutação do DNA mitocondrial é cerca de dez vezes superior à do DNA nuclear. Como a região codificante é quinze vezes maior que a região controlo, a maior parte da variação do DNAm_t ocorre na região codificante (Parsons & Coble, 2001).

O DNA mitocondrial é especialmente útil em Genética Forense devido a características particulares (Thompson & Black, 2007), tais como: (1) elevado número de cópias relativamente ao DNA nuclear (em cada célula estão presentes 1000 -10000 cópias); (2) herança materna e (3) falta de recombinação.

1.2.2.1. Single Nucleotide Polymorphism (SNP)

A variabilidade genética entre os indivíduos é gerada pela sequência de bases presentes no DNA que, apesar de serem apenas quatro, podem combinar-se formando milhões de sequências diferentes. No genoma humano podem ocorrer variações no DNA entre indivíduos que se designam por polimorfismos. Estes são extremamente comuns, aparecendo em pelo menos 1% da população.

Na população humana existem duas principais formas de variação ao nível do DNA: polimorfismos de sequência e polimorfismos de comprimento (Figura 5) (Butler, 2009).

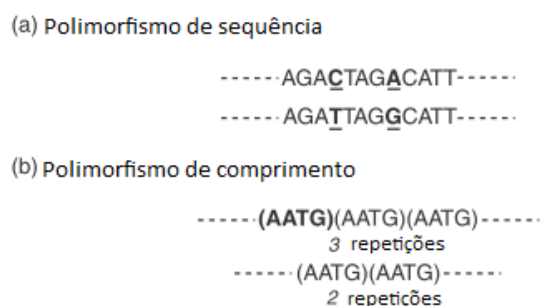


Figura 5 – Variações existentes no DNA:
(a) polimorfismos de sequência e (b) polimorfismos de comprimento.
(Adaptada de (Butler, 2009))

Dos polimorfismos de sequência destacam-se os SNPs (*Single Nucleotide Polymorphisms*) que surgem na sequência de uma mutação pontual, na qual um par de bases é substituído por outro. Os polimorfismos de comprimento – STRs (*Short Tandem*

Repeat) ou microssatélites – são constituídos por sequências de DNA de 2-7 pb que se repete em *tandem* de 5-30 vezes.

Single nucleotide polymorphism, também conhecido como SNP (pronunciado ‘*snip*’), define-se como uma variação estável na sequência de DNA, que ocorre quando um único nucleótido – A, T, C, ou G – no genoma difere entre sequências homólogas (Figura 6). Para a distinção entre SNPs e mutações pontuais, é necessário que o alelo menos comum ocorra com uma frequência de 1% ou mais (Twyman, 2003).

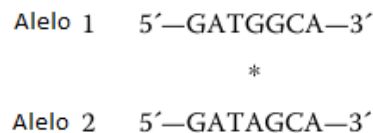


Figura 6 – *Single Nucleotide Polymorphism* (SNP). Representação esquemática de dois alelos que diferem numa posição indicada pela estrela (*). (Adaptado de Thompson and Black (2007))

Estes polimorfismos representam a classe mais simples e abundante de variações no genoma humano (90% de todos os polimorfismos do DNA humano), encontrando-se um em cada 100 a 300 bases. Isto significa que, em todo o genoma humano, existem aproximadamente 10 a 30 milhões potenciais SNPs, apesar de só se encontrarem identificados cerca de 6 milhões (Javed & Mukesh, 2010).

Os SNPs, são na sua maioria, polimorfismos bi-alélicos, mas podem ser tri- e tetra-alélicos. No entanto, nos humanos os SNPs tri- e tetra-alélicos são raros ou quase inexistentes, pois a taxa de mutação num par de bases particular do genoma é extremamente baixa (em média 1 mutação em 100 milhões de gerações), sendo altamente improvável que ocorram duas mutações pontuais na mesma posição do genoma. Por esta razão, os SNPs podem ser usados para distinguir populações (Brookes, 1999; Pinheiro, 2010).

Existem dois tipos de substituições nucleotídicas que resultam em SNPs: 65% das substituições são transições que ocorrem entre purinas (Adenina e Guanina) ou entre pirimidinas (Citocina e Timina), enquanto os 35% restantes são transversões, que ocorrem entre uma purina e uma pirimidina (Martino, Mancuso, & Rossi, 2010).

Os SNPs revolucionaram a genética molecular humana, proporcionando um vasto painel de marcadores genéticos distribuídos por todo o genoma. Atualmente, os SNPs são amplamente investigados em diversas áreas, tais como: estudos da evolução e dinâmica populacional, investigação da genética em doenças complexas, análises forenses e de diagnóstico, entre outros. Estes possuem algumas características que os tornam potenciais marcadores genéticos em estudos forenses (Butler, 2009, 2012; Carracedo, 2005; Sobrino, Brión, & Carracedo, 2005), tais como:

- 1) Poderem ser analisados em *amplicons* curtos, pois o tamanho do produto amplificado é crítico para uma amplificação bem-sucedida de amostras degradadas;
- 2) Possuírem uma taxa de mutação de aproximadamente 100 mil vezes mais baixa que os STRs (10^{-8} vs. 10^{-3}), podendo ser característicos de uma determinada população;
- 3) Poderem ser analisados em *multiplex* (em simultâneo) com um número superior de polimorfismos que os STRs;
- 4) O processamento das amostras e a análise dos resultados poder ser mais automatizada, por não ser necessária a separação baseada no tamanho dos produtos amplificados, como acontece nos loci STR;
- 5) Não existir artefactos *stutter* associado a cada alelo, o que ajuda a simplificar a interpretação do alelo;
- 6) A capacidade de prever a origem étnica e certas características físicas através da seleção cuidadosa de marcadores SNPs.

A análise de polimorfismos no DNA mitocondrial, em particular, as regiões hipervariáveis I e II da região controlo, contêm um poder limitado de discriminação num contexto forense, antropológico e clínico. Uma vez que a sequenciação da molécula mitocondrial não é prática, a análise de informações adicionais da região codificante do genoma mitocondrial (e.g. informações provenientes dos SNPs) provou ser uma boa estratégia para contornar esta desvantagem, embora apresente uma densidade de variabilidade menor que os polimorfismos da região controlo (Alvarez-Iglesias, Jaime, Carracedo, & Salas, 2007; Quintáns et al., 2004).

Ao estudar a região codificante, entra-se numa região em que as mutações podem estar diretamente ligadas a uma variedade de patologias, encontrando-se caracterizadas mais de 250 patologias associadas a mutações do DNAm nesta região. Deste modo, para a investigação forense e genética populacional, estudam-se SNPs localizados nos intervalos, que ocorrem entre os genes mitocondriais, ou nas posições silenciosas dos genes que codificam proteínas, sendo estas posições consideradas não codificantes (Pinheiro, 2010).

1.2.2.2. Filogeografia mitocondrial

Os polimorfismos presentes no DNA mitocondrial, que se encontram distribuídos por todo o genoma mitocondrial, possuem as particularidades de não recombinar (deste modo, são herdados em blocos haplótipos) e a herança, nas gerações seguintes, é estável. Estes aspetos são úteis em estudos de evolução humana e análises forenses, pois as mutações são acumuladas nas moléculas, sendo transferidas de geração em geração como uma impressão digital genética da história evolucionária. A análise dessas variações permite inferir a história da evolução humana e os padrões de migração, estudando as

linhagens presentes (McNevin, Bate, Daniel, & Walsh, 2011; Quintáns et al., 2004; Zuccarelli et al., 2010).

A evolução do genoma mitocondrial humano é caracterizada pelo aparecimento de linhagens monofiléticas, ou haplogrupos, definidos por um conjunto de mutações que ocorreram há milhares de anos. Os haplogrupos mitocondriais foram definidos pela primeira vez, em estudos de genética populacional, através de RFLPs (Torroni et al., 1993; Torroni et al., 1992), permitindo a distribuição geográfica do DNAmT em haplogrupos Europeus, Asiáticos e Africanos, que se ramificaram em subhaplogrupos mitocondriais.

A árvore filogenética do DNAmT humano (Figura 7) deriva de um ancestral comum africano que existiu há cerca de 200.000 anos (Cann, Stoneking, & Wilson, 1987) e encontra-se dividida em dois ramos principais, L0 e L1'2'3'4'5'6 (L1'6).

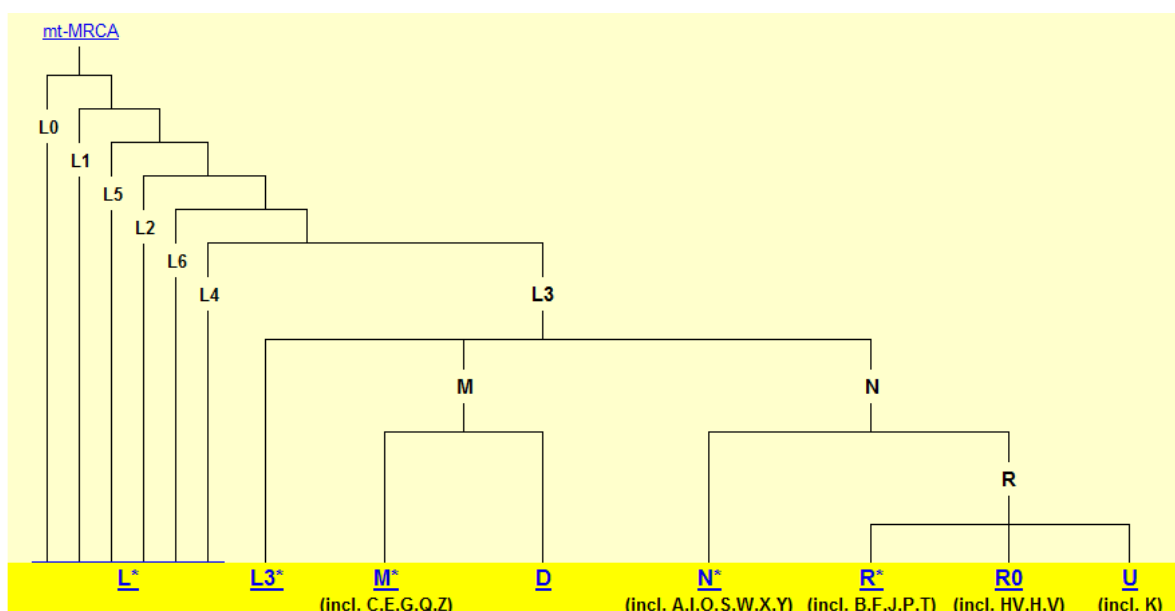


Figura 7 – Representação esquemática da filogenia do DNA mitocondrial.
(adaptado de Mannis vanOven and Kayser (2009))

Os haplogrupos L0, L1, L2 e L3 são os principais haplogrupos africanos com origem na África Sub-Sahariana, sendo estes os mais antigos e diversos. O haplogrupo L0 é considerado o ancestral por se encontrar posicionado na raiz da árvore filogenética mitocondrial.

O haplogrupo L3, que os africanos partilham com o resto do mundo, é composto por dois subhaplogrupos, os macrohaplogrupos M e N, que deixaram a África há cerca de 60.000-80.000 anos (Richards, Bandelt, Kivisild, & Oppenheimer, 2006; Watson, Forster, Richards, & Bandelt, 1997), e que, com a saída do Homem Moderno de África e subsequente colonização da Ásia e da Europa, deram origem às linhagens de DNAmT presentes na restante população, com uma redução drástica de variabilidade do DNAmT.

As linhagens europeias, derivadas do macrohaplogrupo N, caracterizam-se por uma grande variabilidade, sendo os haplogrupos H, I, J, K, T, U, V, W e X representativos de mais de 98% desta população. O haplogrupo H é o mais representativo, que compreende cerca de 45% de todas as linhagens mitocondriais europeias.

Os haplogrupos C, D, E, G e Z, derivados do macrohaplogrupo M, assim como os haplogrupos A, B, F e Y, derivados dos macrohaplogrupo N, são os mais comuns nas linhagens com características asiáticas. Entre os nativos Americanos, estão presentes os haplogrupos asiáticos A, B, C e D (Pinheiro, 2010; Salas, Bandelt, Macaulay, & Richards, 2007).

1.3. Metodologias de Genotipagem de SNPs

Nos últimos anos o interesse nos SNPs aumentou significativamente devido à sua abundância no genoma humano e, simultaneamente, às novas tecnologias de detecção de SNPs que têm vindo a ser desenvolvidas para localizar e caracterizar variações num local específico do genoma humano (Kwok, 2001; Sobrino et al., 2005).

A escolha do método apropriado para genotipagem de SNPs é crítica para o sucesso de um estudo e depende da elevada sensibilidade, precisão e reprodutibilidade do método; assim como da capacidade de genotipar vários loci em simultâneo (sistema *multiplex*) e o nível de rendimento (número de marcadores genotipados em cada ciclo) (Sobrino et al., 2005).

As metodologias de genotipagem de SNPs consistem na combinação de duas fases diferentes: uma destina-se a determinar o tipo de base presente num dado *locus* (discriminação alélica), e a outra a verificar a presença de alelos (detecção de sinal). Os produtos resultantes das reações de discriminação alélica podem ser detetados usando vários métodos, sendo que o método de detecção (sequenciador automático) pode analisar produtos resultantes de diferentes metodologias.

Segundo Twyman (2005), existem três métodos gerais de discriminação alélica: (1) extensão de *primer* (minissequenciação), (2) hibridização/*annealing* e (3) separação enzimática. Para a análise dos produtos de cada tipo de reação, existem vários métodos de detecção, tais como: fluorescência, luminescência, espectrometria de massa, entre outros.

Os métodos que se baseiam na extensão ou hibridação de *primers* são utilizados em estudos populacionais e aplicações forenses e serão descritos nos próximos pontos:

1.3.1. Minissequenciação – Extensão de *Primer*

A extensão de *primer* baseia-se na capacidade da DNA polimerase incorporar dideoxynucleotídeos (ddNTPs) específicos complementares com a sequência de DNA alvo. Esta metodologia, também designada por minissequenciação ou extensão de base única (*Single base extension* – SBE), é um método de discriminação alélica robusto, fiável, rápido e de baixo custo. A capacidade de, numa única reação, tipar simultaneamente diversos SNPs, tornou a minissequenciação uma das técnicas mais aceite para aplicações forenses (Kwok, 2001).

A base de reação do método de minissequenciação é muito similar à sequenciação de Sanger (Anderson et al., 1981). Este método, primeiro descrito por Tully et al. (1996), implica a amplificação da região alvo a partir de DNA genômico e posterior purificação do DNA amplificado. De seguida, um *primer* não marcado hibrida com o produto de amplificação na base imediatamente adjacente ao local da existência do SNP. Posteriormente, na presença de quatro ddNTPs marcados com diferentes fluoróforos, ocorre a extensão de uma única base, como se encontra esquematizado na Figura 8. A extensão do *primer* pode ser analisada utilizando a eletroforese capilar e a cor e tamanho do pico detetado permite caracterizar o SNP presente (Butler, 2012; Goodwin, Linacre, & Hadi, 2011; Sobrino et al., 2005).

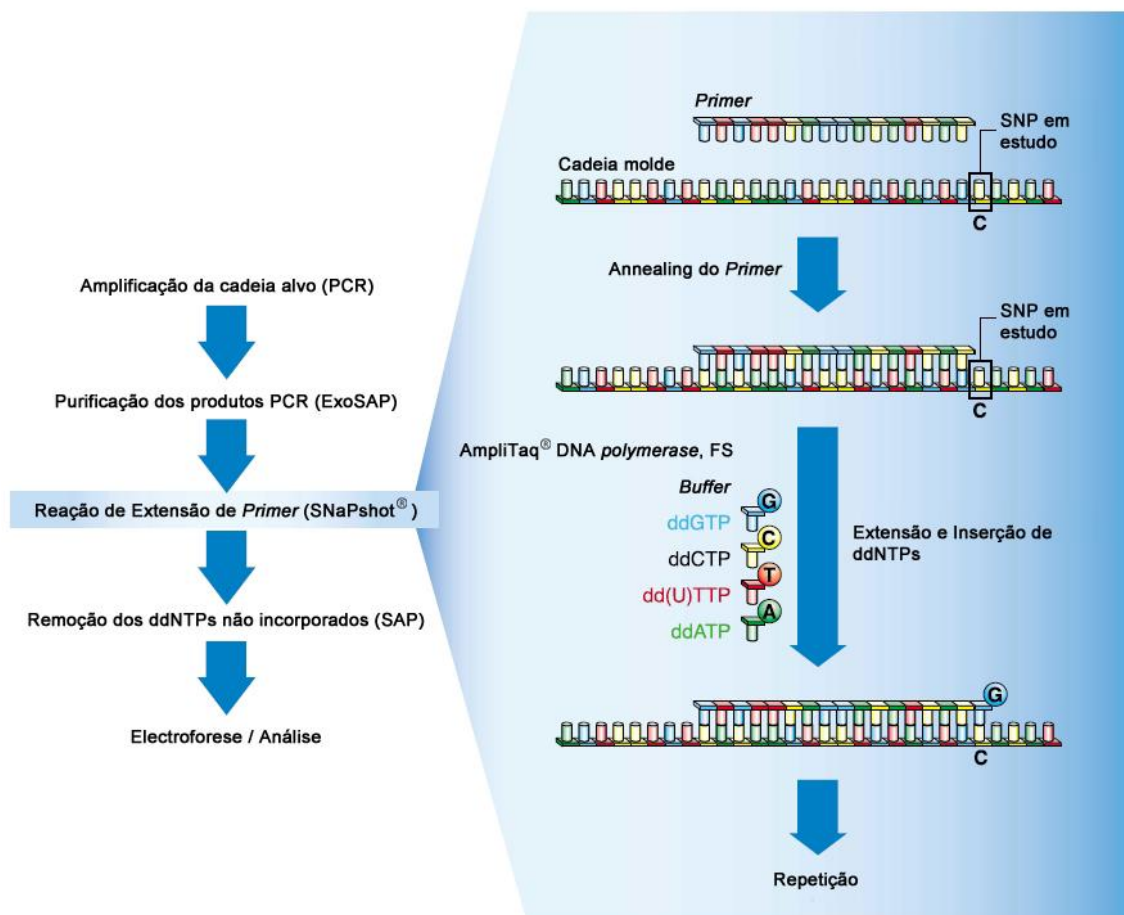


Figura 8 – Esquema representativo do método de minissequenciação. (Adaptado de Biosystems (2001)).

Este método utiliza o kit comercial SNaPshot™ Multiplex Kit (Applied Biosystems) que tem por base o método de minissequenciação, seguida de eletroforese e deteção de fluorescência.

Ao longo dos últimos anos, a técnica de minissequenciação tem vindo a ser estudada na região codificante do DNAm por vários autores em diversas populações (Brandstätter, Parsons, & Parson, 2003; Quintáns et al., 2004; Umetsu et al., 2005). A popularidade desta técnica para a genotipagem de SNPs deve-se às diversas vantagens que esta apresenta, tais como: (1) o poder de discriminação em estudos mitocondriais, pode aumentar através de

um *design* sólido baseado em critérios filogenéticos; (2) permite desenho *multiplex*, incluindo uma grande quantidade de SNPs localizados em locais distantes do DNAm; (3) é um método rápido e de baixo custo (comparativamente a outras técnicas de genotipagem de SNPs); (4) genotipagem em *multiplex* previne recombinações artefactuais, que facilmente ocorrem quando se genotipa um SNP de cada vez (Mosquera-Miguel et al., 2009).

1.3.2. PCR em Tempo Real

A PCR em Tempo Real – também designada de PCR quantitativa (Q-PCR, qPCR, qrt-PCR) – foi pela primeira vez descrita por Russell Higuchi e seus colaboradores no início dos anos 90 (Higuchi, Fockler, Dollinger, & Watson, 1993). Higuchi e seus colaboradores construíram um sistema ao qual acoplaram uma câmara de vídeo, de modo a monitorizar a PCR a totalidade dos ciclos. Este mecanismo permitia-lhes detetar o aumento da fluorescência durante a reação, devido à ligação de brometo de etídio às moléculas de DNA de dupla cadeia recém-sintetizadas.

Esta técnica baseia-se no processo de PCR convencional diferindo principalmente numa característica inovadora: a possibilidade da quantificação em tempo real do DNA amplificado em cada ciclo. A deteção dos produtos de amplificação é realizada através da emissão de fluorescência de fluoróforos que são incluídos na reação, sendo a quantidade de fluorescência emitida diretamente proporcional à quantidade do produto de PCR (Figura 9). Neste método, as fases de amplificação, deteção e quantificação são totalmente automatizadas, ocorrendo em simultâneo e em tempo real (Heid, Stevens, Livak, & Williams, 1996).

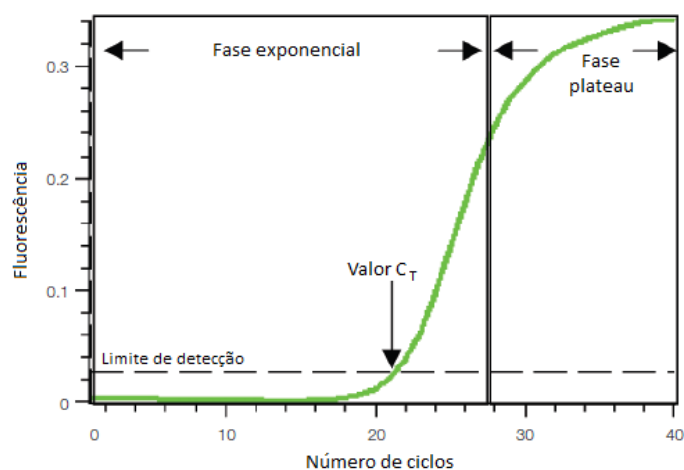


Figura 9 – Gráfico de amplificação da PCR em Tempo Real.
(Adaptado de Bio-Rad Laboratories (2006))

Os métodos químicos de fluorescência utilizados na PCR em Tempo Real são diversificados, recorrendo todos eles a compostos fluorescentes cujo sinal é suscetível de ser detetado por um laser do termociclador ao longo do processo. Os fluoróforos podem ser divididos em dois grandes grupos, corantes intercalares e sondas de sequência

específica, de acordo com o tipo de composto fluorescente e respetivo comportamento durante o processo (Bio-Rad Laboratories, 2006; Mackay, 2007).

O método de corantes intercalares (e.g. SYBR® Green, LCGreen® Plus, EvaGreen® e SYTO® 9) é a forma mais simples e menos dispendiosa de monitorizar a PCR em Tempo Real. Estes corantes apresentam fluorescência apenas quando se ligam a DNA de dupla cadeia. Deste modo, à medida que o número de cópias de DNA aumenta durante a reação, a fluorescência também aumenta (Figura 10 – A). A maior desvantagem da utilização deste método é a falta de especificidade, pois estes corantes ligam-se a qualquer DNA de cadeia dupla, incluindo dímeros de primers e outros produtos não específicos. Uma vez que este método não é específico para uma determinada sequência de DNA, este só pode ser utilizado em reações *Singleplex*.

O método de sondas de sequência específica utiliza sondas fluorescentes, como por exemplo: sondas *TaqMan*®, sondas de hibridação *FRET*, sondas *primers* ‘Scorpion’ e ‘Sunrise’, que são porções de DNA com especificidade para uma determinada sequência de DNA, marcadas com fluorocromos. Este método apresenta duas grandes vantagens em relação aos corantes intercalares: (1) detetam especificamente a sequência alvo, deste modo, os produtos não específicos não são detetados; (2) permitem efetuar reações *Multiplex* (Bio-Rad Laboratories, 2006).

As sondas *TaqMan* são o tipo de sondas mais simples e usualmente utilizadas para discriminação alélica. Estas sondas apresentam na extremidade 5’ uma molécula fluorescente denominada *reporter* (R) e na extremidade 3’ outra molécula fluorescente designada *quencher* (Q), encontrando-se o *reporter* próximo do *quencher*. Enquanto a sonda se encontra intacta a emissão de fluorescência é muito baixa ou nenhuma devido à capacidade do *quencher* absorver a fluorescência do *reporter*. Durante a PCR, a sonda liga-se à sequência alvo e na extensão o *reporter* é clivado pela atividade 5’ da Taq DNA polimerase. Esta clivagem separa o *reporter* do *quencher*, permitindo o aumento da fluorescência (Figura 10 – B) (Butler, 2012; Goodwin et al., 2011).

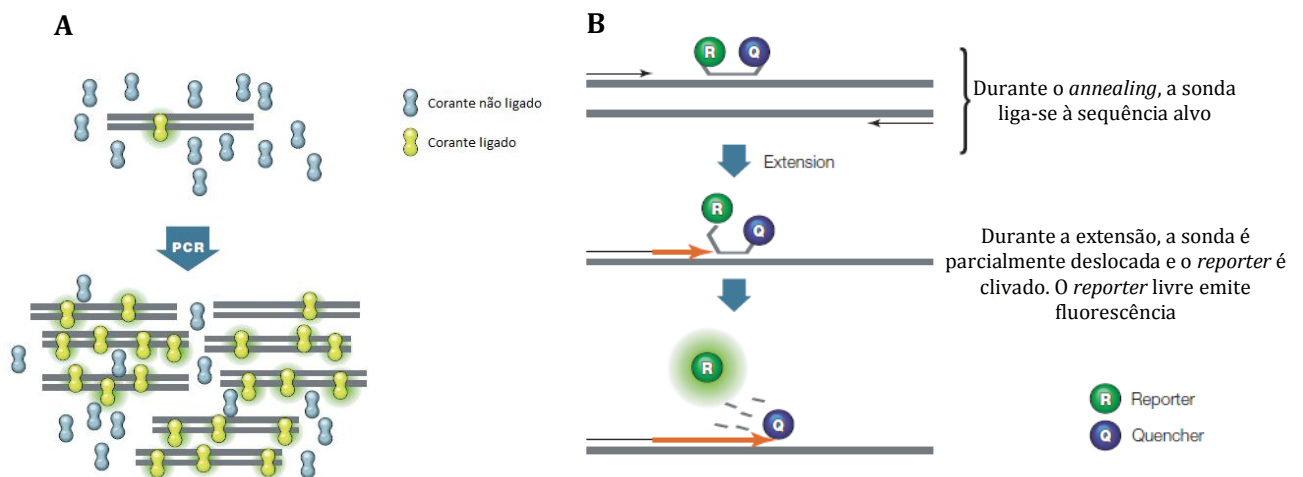


Figura 10 – Esquema ilustrativo de dois grandes métodos químicos de fluorescência.
A - Corante intercalares; B - Sonda de sequência específica
(Adaptado de Bio-Rad Laboratories (2006))

Análise HRM (*High-Resolution Melting Analysis*)

Uma das abordagens mais utilizadas para a análise dos resultados obtidos da PCR em Tempo Real é a análise de curvas de *melting*. A desnaturação das cadeias de DNA pode ser seguida com corantes intercalares, que apresentam fluorescência na presença de DNA de cadeia dupla. A fluorescência do DNA é continuamente monitorizada, dependendo da temperatura (Figura 11), e pode ser analisada com base na temperatura de melting (T_m) – temperatura em que 50% do DNA se apresenta em dupla cadeia e o restante em cadeia simples - ou no perfil da curva de *melting*. O perfil (forma) da curva de *melting* depende do conteúdo em GC, tamanho e composição dos nucleótidos do produto amplificado (Erali, Voelkerding, & Wittwer, 2008; Farrar, Reed, & Wittwer, 2010).

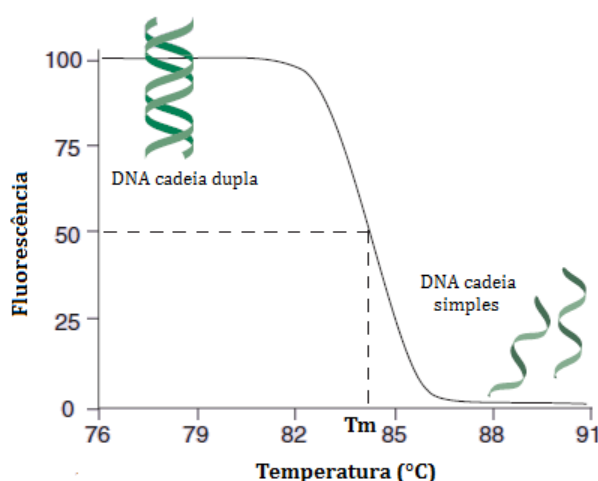


Figura 11 – Figura ilustrativa de uma curva de *Melting*.
(Adaptado de Reed, Kent, and Wittwer (2007))

A análise das curvas de *melting* em conjunto com a PCR em Tempo Real foi introduzida em 1997 (Wittwer, Herrmann, Moss, & Rasmussen, 1997; C Wittwer et al., 1997) com a utilização do LightCycler®. Vários anos mais tarde, foi introduzida uma nova técnica para a análise de variações genéticas no produto de PCR, designada por *High-Resolution Melting Analysis* (também conhecido por análise HRM ou HRMA) (Wittwer, Reed, Gundry, Vandersteen, & Pryor, 2003).

High-Resolution Melting é uma análise pós-PCR, utilizada para identificar variações genéticas nas sequências de DNA, que se baseia na análise clássica das curvas de *melting*. A análise HRM tornou-se uma técnica interessante para a genotipagem de SNPs, devido à sua simplicidade, baixo custo, fácil utilização, e elevada sensibilidade, sendo que quanto menor forem os fragmentos (80-100 pb) maior é a sensibilidade. (Vossen, Aten, Roos, & Dunnen, 2009; Wittwer, 2009).

A análise HRM foi descrita pela primeira vez utilizando sondas marcadas com fluorescência para distinguir diferentes heterozigóticos e homozigóticos (Gundry et al., 2003). Contudo, este método apenas detetava variações na região de hibridação da sonda, e todas as variações fora dessa região não eram detetadas. As sondas foram rapidamente

substituídas por corantes, eliminando qualquer necessidade de utilizar sondas com marcação fluorescente (Wittwer et al., 2003). O primeiro corante desenvolvido para a genotipagem foi o SYBR® Green I, mas este apresentava várias limitações, tais como, a redistribuição do corante durante a dissociação das cadeias de DNA (Figura 12 – A) e a inibição da PCR quando este era utilizado em elevadas concentrações. Mais tarde foram desenvolvidos corantes saturados, tais como: LCGreen®, Syto 9® e EvaGreen®, que saturavam todo o DNA de dupla cadeia produzido durante a PCR, eliminando a redistribuição do corante (Figura 12 – B), sem inibir a PCR (Dorak, 2007).

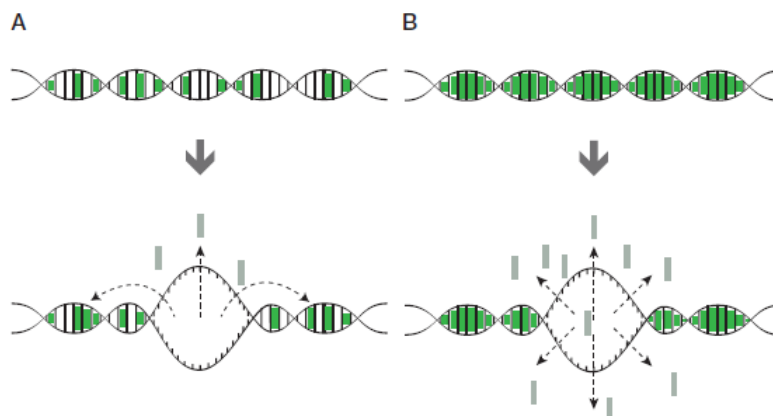


Figura 12 – Esquema ilustrativo do modo de atuação dos corantes não saturados (A) e corantes saturados (B) durante a dissociação das cadeias de DNA. (Adaptado de Taylor et al. (2010))

Os resultados obtidos através da análise HRM podem fornecer informações importantes para detecção de mutações, genotipagem de SNPs, análise de metilações de DNA, entre outras. A utilização desta análise para o estudo do DNA mitocondrial tem sido alvo de poucos estudos. Gidlöf e colegas (2009) desenvolveram um ensaio para a discriminação de diferentes indivíduos, baseado na análise HRM das regiões HVS I e HVS II. Estes autores demonstraram que o referido ensaio pode funcionar como um método rápido e pouco dispendioso de pré-seleção das amostras antes da sequenciação de DNA. Por sua vez, Dobrowolski e seus colegas (2009) utilizaram a análise HRM para identificar mutações patológicas na sequência do DNA mitocondrial.

O reduzido número de estudos efetuados recorrendo à análise HRM poderá estar correlacionado com o facto desta análise não poder ser facilmente aplicada em *Multiplex*, como sugerido por Vossen et al. (2009).

1.3.3. Comparação das metodologias: Vantagens e Desvantagens

Nesta secção irão ser resumidas as principais vantagens e desvantagens das metodologias abordadas anteriormente: Minissequenciação e PCR em Tempo Real (Análise HRM) (Tabela 1).

Tabela 1 – Tabela comparativa das metodologias: Minissequenciação e Análise HRM.

	Minissequenciação	PCR em Tempo Real (Análise HRM)
Vantagens	<ul style="list-style-type: none"> • Método robusto, sensível e fiável; • Capacidade de amplificação simultânea de vários <i>loci</i>; • Análise de produtos de amplificação pequenos (< 50bp). 	<ul style="list-style-type: none"> • Simples, rápido, de fácil utilização e pouco dispendioso; • Procedimento num só passo.
Desvantagens	<ul style="list-style-type: none"> • Método com procedimento mais complexo – elevado tempo de resposta. 	<ul style="list-style-type: none"> • Não fornece informações sobre a mutação específica presente no <i>amplicon</i>; • Tamanho do <i>amplicon</i>. • Não é facilmente aplicada em <i>multiplex</i>; • Depende de um bom PCR, instrumento e corante.

A minissequenciação é um dos métodos mais utilizados e estudados para a genotipagem de SNPs devido às suas especificidades. Uma das suas principais vantagens é a capacidade de realizar ensaios *multiplex*, ou seja, numa só reação consegue-se genotipar diversos SNPs. Como o poder de discriminação dos SNPs é muito inferior à dos STRs, a utilização de SNPs em Genética Forense torna-se mais alcançável com um método que caracterize diversos SNPs numa só reação. Em Genética Forense, por vezes, o DNA encontra-se parcialmente degradado, impossibilitando a genotipagem de STRs, podendo a genotipagem de SNPs ser a única alternativa uma vez que fragmentos de DNA inferiores a 50 pb podem ser analisados recorrendo a este método (Mosquera-Miguel et al., 2009).

Para além destas vantagens, a minissequenciação também possui a vantagem de ser um método robusto, sensível e fiável, pois este método utiliza *primers* específicos que hibridam com o produto de amplificação na base imediatamente antes do local do SNP em estudo, podendo-se determinar com precisão qual o polimorfismo presente (Mosquera-Miguel et al., 2009).

No entanto, a minissequenciação possui a desvantagem de ser um método mais complexo, o que resulta numa abordagem com vários passos quando comparado com a

análise HRM. Para a obtenção de resultados na minissequenciação é necessário efetuar vários procedimentos, desde a amplificação do DNA, purificação pós-PCR, minissequenciação com o kit SNaPshot™ e purificação pós-SNaPshot® e, finalmente a separação e detecção dos fragmentos obtidos.

Por sua vez, a análise HRM é uma técnica recentemente utilizada para identificar variações genéticas nas sequências de DNA e possui como principais vantagens a simplicidade, rapidez e fácil utilização. Este método é processado num único passo, ou seja, não é necessário processamento da amostra entre a amplificação e a análise dos produtos amplificados, diminuindo a contaminação e tornando a análise pouco dispendioso em comparação com a minissequenciação (Gidlöf et al., 2009; Wittwer, 2009).

No entanto, a análise HRM possui como principal desvantagem o facto de não fornecer informações sobre a mutação específica presente no *amplicon*, sendo necessário o uso de amostras controlo, cujos resultados para os marcadores moleculares em estudo são conhecidos.

O tamanho do *amplicon* pode influenciar a sensibilidade e especificidade da análise HRM. O uso de *amplicons* de pequeno tamanho possui vantagens e desvantagens. Como vantagens: os *primers* podem ser selecionados o mais próximo possível do SNP, sendo os produtos de PCR de tamanho mais reduzido o que permite uma diferenciação superior das curvas de *melting*; e como desvantagens: a escolha dos *primers* ser menos flexível e a distinção entre dímeros de *primers* e produtos de amplificação ser mais difícil. Acresce que, à medida que o tamanho dos *amplicons* aumenta, é possível a detecção de SNPs adicionais o que resulta num perfil de *melting* mais complexo, sendo a distinção entre variações de sequências mais difícil. Deste modo, *amplicons* com tamanhos compreendidos entre os 80-100 pb são os ideais para utilizar em Genética Forense (Taylor et al., 2010).

Este método também possui a desvantagem de não ser facilmente aplicado em *multiplex*, ou seja, a determinação de dois ou mais polimorfismos no mesmo *amplicon* pode ser difícil. Desta forma, é fundamental a utilização de *primers* que amplifiquem apenas um SNP, embora diferentes mutações no mesmo *amplicon* forneçam curvas de *melting* diferentes.

Na análise HRM, a utilização de amostras e controlos com concentrações de DNA similares, um bom equipamento, e corante são essências para a obtenção de um bom produto de PCR. A amplificação não específica dos produtos gerados durante a PCR pode levar a um comportamento diferente das curvas *melting*, afetando a qualidade dos resultados finais.

2. OBJETIVOS

No presente trabalho foi estudada uma amostra populacional de indivíduos naturais de Timor-Leste, recorrendo a marcadores bi-alélicos localizados no DNA mitocondrial, tendo como objetivos principais:

1. Comparar duas metodologias de genotipagem de SNPs: (i) Minissequenciação, com a utilização do kit SNaPshot™ *Multiplex*, e (ii) PCR em Tempo Real, através da análise HRM;
2. Comparar e discutir as vantagens e limitações das metodologias aplicadas com interesse forense.
3. Determinar os haplogrupos presentes na amostra, de acordo com polimorfismos das regiões HVS I e HVS II e região codificante do DNA mitocondrial;

A escolha do método apropriado para genotipagem de SNPs é importante para o sucesso de um estudo e depende da sensibilidade, precisão e reprodutibilidade do método. Com a realização destes objetivos, pretende-se confrontar duas metodologias de genotipagens de SNPs, atualmente muito utilizadas, com possível utilidade na resolução de perícias médico-legais no âmbito da Genética e Biologia Forense. Pretende-se igualmente efetuar um estudo populacional, a fim de proporcionar uma ideia aproximada das frequências dos SNPs estudados na amostragem populacional de Timor-Leste, tendo em consideração o reduzido número de amostras.

3. MATERIAL E MÉTODOS

3.1. Amostragem populacional

Neste trabalho foram estudadas 38 amostras de indivíduos naturais de Timor-Leste.

As amostras utilizadas pertencem à coleção de Timor-Leste do Laboratório de Genética Aplicada do Departamento de Biologia da Universidade de Aveiro. A origem e grupo etnolinguístico dos indivíduos intervenientes, bem como dos seus pais e avós, foram obtidos através da realização de inquéritos, durante a recolha das amostras. A recolha das amostras foi voluntária e obedeceu aos critérios básicos de consentimento informado.

O material biológico utilizado foi sangue venoso, colhido para tubos *eppendorf* de 1,5 mL. As amostras foram armazenadas a -20°C até se proceder à extração de DNA, a partir do sangue total.

3.2. Extração de DNA

A extração do DNA foi efetuada recorrendo ao método modificado de *Chelex*® 100 (Walsh, Metzger, & Higuchi, 1991), descrito no Anexo I. O *Chelex*® 100 é uma resina composta por copolímeros estireno-divinilbenzeno, com grande afinidade para iões metálicos polivalentes, tais como o magnésio (Mg^{2+}), prevenindo assim a degradação do DNA em presença de iões metálicos (Goodwin et al., 2011).

A extração por resina de *Chelex* constitui um método simples e rápido de obtenção de DNA, não envolvendo solventes orgânicos nem transferências múltiplas de tubos.

Após extração procedeu-se à quantificação do DNA, utilizando para tal o Qubit® Fluorometer (Invitrogen™). A quantidade de DNA utilizada para este trabalho foi de 5 ng.

3.3. Genotipagem de SNPs

No presente trabalho foram estudados 32 SNPs, que se encontram indicados a vermelho na árvore filogenética do DNAmT (Figura 13), descritos por Alvarez-Iglesias et al. (2007). Os SNPs selecionados definem 39 haplogrupos, que se encontram localizados nos principais ramos da árvore filogenética do DNA mitocondrial, característicos da Ásia (VanOven & Kayser, 2009).

Para a caracterização dos SNPs em estudo, utilizou-se duas metodologias: Minissequenciação, através do kit SNaPshot™ *Multiplex* e PCR em Tempo Real, recorrendo à análise HRM.

3.3.1. Minissequenciação

Com esta metodologia estudou-se 32 SNPs, que foram divididos em dois *multiplex*. A necessidade de dividir os SNPs em duas reações *multiplex* prendeu-se com a limitação de deteção de um elevado número de marcadores ao utilizar o kit SNaPshot™ *multiplex*. A identificação dos SNPs, a sequência dos *primers* e as concentrações finais na reação encontram-se descritas na Tabela 2.

Cada reação *multiplex* foi realizada num volume final de 15 µl, contendo 7,5 µl de 2x QIAGEN *Multiplex* PCR Master Mix (Qiagen) a 1x; 2 µl de *primer* mix (consoante o *multiplex*) preparado de acordo com as concentrações propostas por Alvarez-Iglesias et al. (2007) e 1 µl de DNA com uma concentração, aproximada, de 5 ng. A amplificação foi levada a cabo num termociclador Icycler (Bio-Rad), utilizando as condições descritas no manual do *kit* (Anexo II).

I) Purificação pós-PCR

Após a confirmação da amplificação através da corrida dos produtos de PCR em gel de agarose 2%, procedeu-se à purificação dos produtos de PCR utilizando a ExoSAP-IT® (USB® Products).

A ExoSAP-IT® é formada por duas enzimas hidrolíticas, a Exonuclease I e a *Shrimp Alkaline Phosphatase* (SAP), que são utilizadas em conjunto num tampão especialmente preparado para remover o excesso de primers e os dNTPs não incorporados. A Exonuclease I degrada *primers* de cadeia simples e qualquer DNA indesejado de cadeia simples, que tenha sido produzido durante a reação de PCR. A SAP remove os dNTPs remanescentes que permanecem na mistura de PCR após a amplificação.

Seguindo as recomendações do fabricante, foram utilizados 2 µl de ExoSAP-IT® para cada 5 µl de produto de PCR, seguindo-se um período de incubação de 37°C, temperatura que permite a atuação das enzimas, durante 15 minutos e um período de 15 minutos a 80°C para a sua inativação.

Tabela 2 - Sequência dos *primers* utilizados para a amplificação de SNPs.

SNP	Primer 5'-3' Fw/Re	Tamanho (pb)	Concentração Final (μM)
Multiplex I			
1119	GCTAAGACCCAACTGGGATT GTTCTGGCGAGCAGTTTTGT	84	0.5
1719/1736	CCAAACCCACTCCACCTTACT GCGCCAGGTTTCAATTCTA	87	0.2
2092	TGCCCACAGAACCCTCTAAA TGTCCAAAGAGCTGTTCTCT	69	0.5
3547/3552	CATCTACCATCACCTCTACATCA ACCAGGGGGTTGGGTATG	93	0.4
3970/4071	GGCTTCAACATCGAATACGC GAAGTAGGGTCTTGGTGACAAAA	168	0.4
5417	CCTCAATCACACTACTCCCATA GGGGTGGGTTTTGTATGTTC	72	0.5
8281-8289del	TAGGGCCCGTATTTACCCTAT AAGAGGTGTTGGTTCTCTTAATCTTT	109	0.5
8414	TACTACCGTATGGCCACCA GGGAGGTAGGTGGTAGTTTGTG	96	0.2
9824	TTTTGTAGCCACAGGCTTCC GAGGAAAGTTGAGCCAATAATGA	53	0.4
12705/12714/12771	CCCAAACATTAATCAGTCTTCAA CTCGGGCGTATCATCAACTG	151	0.4
13759	GGAAGCCTATTCGCAGGATT GTTGTTTGAAGGGGGATG	64	0.4
Multiplex II			
4491	ATTAATCCCCTGGCCCAAC GATGAGTGTGCCTGCAAGA	56	0.2
4820/4824/4883	TAGCCCCCTTTCACTTCTGA GGCTTACGTTTAGTGAGGGAGA	135	0.4
10310/10398/10400	CCATGAGCCCTACAAACAAC TGAGTCGAAATCATTCGTTTTG	158	0.5
11959/11969/12007	AACCACGTTCTCCTGATCAAA AATGTGGTGGGTGAGTGAGC	120	0.5
12338/12358	TGGTCTTAGGCCCCAAAAAT GGGGAATTAGGGAAGTCAGG	97	0.4
14502/14569	GACAACCATCATTCCTCCCTA CTATTTATGGGGGTTTAGTATTGATT	124	0.4
15487/15535	CGGCTTACTTCTCTTCTCTCT ATATCATTCGGGCTTGATGTG	129	0.5

II) Minissequenciação com o kit SNaPshot™ Multiplex

A reação de minissequenciação foi realizada utilizando o kit ABI Prism® SNaPshot™ Multiplex (Applied Biosystems). Deste kit fazem parte ddNTPs marcados com diferentes fluoróforos e os restantes reagentes, bem como os respetivos *primers* não marcados.

As sequências dos *primers* e correspondentes concentrações finais na reação encontram-se descritas na Tabela 3.

Cada reação *multiplex* foi realizada num volume final de 10 µl, contendo 3,5 µl de SNaPshot Multiplex Ready Reaction Mix; 1,5 µl de mix de *primers* (consoante o *multiplex*) preparado de acordo com as concentrações propostas por Alvarez-Iglesias et al. (2007) e 1,5 µl de produto de PCR purificado. A reação foi levada a cabo num termociclador Icyler (Bio-Rad) seguindo as condições do fabricante (Anexo III).

III) Purificação pós-SNaPshot™

Após a minissequenciação procedeu-se à remoção dos ddNTPs não incorporados na reação, através da utilização da SAP (*Shrimp Alkaline Phosphatase*) (Promega). A SAP desfosforila todos os ddNTPs remanescentes.

De forma a purificar o produto da minissequenciação, foi adicionado 1 µl de SAP aos 10 µl de produto minissequenciado, seguindo-se um período de incubação a 37°C, temperatura que permite a atuação da enzima, durante 60 minutos e um período de 15 minutos a 80°C para a sua inativação

Tabela 3 - Sequência dos *primers* utilizados para a reação de minissequenciação.

SNP	Extensão do <i>primer</i> ¹	Tamanho (pb)	Polimorfismo	Cadeia ²	Concentração Final (μM)
Multiplex I					
1119	(<i>gact</i>) ₈ GTTAGCCCTAAACCTCAACAGTTAAA	58	T→C	L	0.40
1719	GACTGACACTCCACCTTACTACCAGACAACCTTA	34	G→A	L	0.15
1736	CCTTAGCCAAACCATTTACCCA	22	A→G	L	0.40
2092	(<i>gact</i>) ₄ ACCCTCTAAATCCCCTTGTAATTTAA	43	C→T	L	0.50
3547	CCCGACCTTAGCTCTCACC	19	A→G	L	0.15
3552A	(<i>gact</i>) ₁₀ GACCTTAGCTCTCACCATCGC	61	T→A	L	0.40
3970	(<i>gact</i>) ₆ GAGTTTGTGTATTCGGCTATGAAGAATA	52	C→T	H	0.20
4071	GACTGACCGCACTCTCCCCTGAACTCTA	28	C→T	L	0.20
5417	(<i>gact</i>) ₉ CCATATCTAACACGTAAAAATAAATGACA	67	G→A	L	0.40
8281-8289del	(<i>gact</i>) ₈ GACGTATTTACCCTATAGCACCCCTCTA	61	C→G ³	L	0.30
8414	(<i>gact</i>) ₃ GACGCCACCATAATTACCCCATATA	37	C→T	L	0.20
9824	(<i>gact</i>) ₃ GCACAGGCTTCACGGACT	31	T→C	L	0.20
12705	(<i>c</i>) ₁₇ AACATTAATCAGTTCTTCAAATATCTACTCAT	49	T→C	L	0.30
12714	(<i>gact</i>) ₅ CAGTTCTTCAAATATCTACTCATCTTCCTAAT	52	T→C	L	0.50
12771	(<i>gact</i>) ₂ CAACTGTTTCATCGGCTGAGA	28	G→A	L	0.15
13759	(<i>gact</i>) ₄ GATCATTACTAACAAACATTTCCCC	40	G→A	L	0.15
Multiplex II					
4491	CCTGCAAAGATGGTAGAGTAGATGA	25	G→A	H	0.20
4820	GAGGGGTGCCTTGGGTAAC	19	G→A	H	0.40
4824	CTTTCACCTCTGAGTCCCAGAGGTT	25	A→G	L	0.20
4883	TCACATGACAAAACTAGCCCC	22	C→T	L	0.20
10310	(<i>gact</i>) ₇ GACATGAGCCCTACAAACAATAACCT	55	G→A	L	0.20
10398	(<i>gact</i>) ₇ TATGAGTGACTACAAAAAGGATTAGACTGA	58	A→G	L	0.40
10400	(<i>c</i>) ₁₁ GTTTTGTTTAACTATATACCAATTC	37	C→T	H	0.40
11959	(<i>gact</i>) ₅ GTCACCTCTCTACTTACAGGACTCAACAT	49	A→G	L	0.30
11969	(<i>gact</i>) ₄ CTACTTACAGGACTCAACATACTAGTCACA	46	G→A	L	0.40
12007	(<i>gact</i>) ₄ CCTCTACATATTTACCACAACACAATG	43	G→A	L	0.15
12338	(<i>gact</i>) ₈ GGTGCAACTCCAAATAAAAGTAA	55	T→C	L	0.50
12358	(<i>gact</i>) ₆ GAGGAAGTCAGGGTTAGGGTGG	46	A→G	H	0.30
14502	(<i>gact</i>) ₄ GAACCATCATTTCCCCTAAATAAA	40	T→C	L	0.50
14569	(<i>gact</i>) ₂ GATAATAACACACCCGACCACACC	32	G→A	L	0.15
15487T	(<i>gact</i>) ₁₁ GCTGGGTGCGCTAGGAGGTC	64	A→T	H	0.40
15535	(<i>gact</i>) ₂ GATTATACCCTAGCCAACCCCTTAAA	34	C→T	L	0.30

¹ As bases em itálico e letra minúscula correspondem às extensões adicionais aos *primers*.

² Cadeia: refere-se à cadeia de DNA alvo na qual o *primer* se liga.

³ A deleção de 9 pares de bases (8281-8289del) é representada como uma mudança de C para G.

IV) Eletroforese Capilar

Após a amplificação e minissequenciação, seguiu-se o passo da detecção. A separação dos fragmentos de SNPs é conseguida através da eletroforese capilar com detecção de fluorescência dos dideoxinucleotídeos marcados com fluoróforos que foram incorporados na minissequenciação (A: dR6G – verde; C: dTAMRA™ - amarelo; G: dR110 – azul; T: dROX™ – vermelho).

Uma vez obtido o produto de minissequenciação purificado, preparou-se as amostras para aplicação no sequenciador automático, adicionando: Formamida desionizada (Hi-Di™ Formamide da Applied Biosystems), GS120-Liz *size standard* e o produto da minissequenciação. De seguida as amostras foram desnaturadas durante 5 minutos a 95°C e depois colocadas no gelo outros 5 minutos.

A aplicação das amostras foi efetuada no sequenciador automático ABI Prism® 310 Genetic Analyzer (Applied Biosystems), utilizando o POP-6™ (*Performance Optimized Polymer*) como polímero. Acresce que para algumas amostras foi testado o POP-4™.

Após a corrida, os resultados foram analisados com recurso ao *software GeneMapper® ID Software v3.2* (Applied Biosystems).

3.3.2. PCR em Tempo Real – Análise HRM

A amplificação do DNA por PCR em Tempo Real foi realizada em *Singleplex* para 9 SNPs dos 32 estudados, descritos na Tabela 4, recorrendo a um kit de amplificação apropriado para a genotipagem de SNP através da análise HRM. Foram testados três kits de amplificação: SsoFast™ EvaGreen® Supermix (Bio-Rad), AccuMelt™ HRM SuperMix (Quanta Biosciences™) e Type-it® HRM™ PCR (Qiagen) de forma a determinar qual o melhor kit, com a utilização do marcador molecular 8281-8289del (deleção de 9pb) e das amostras: controlo negativo, amostra controlo e duas amostras das 38 objeto de estudo (amostras 1 e 45). A plataforma de PCR utilizada foi *CFX96™ Real-Time PCR* (Bio-Rad), sendo a análise HRM efetuada com o *Software Precision Melt Analysis™*.

Tabela 4 – Sequência dos *primers* utilizados para a amplificação e detecção de SNPs na PCR em Tempo Real.

SNP	Polimorfismo	Primer 5'-3' Fw/Re	Tamanho (pb)	Referência
1119	T→C	GCTAAGACCCAACTGGGATT GTTCTGGCGAGCAGTTTTGT	84	Alvarez-Iglesias et al. (2007)
2092	C→T	TGCCCACAGAACCCTCTAAA TGTCCAAAGAGCTGTTCTCT	69	Alvarez-Iglesias et al. (2007)
4491	C→T	ATTAATCCCCTGGCCCAAC GATGAGTGTGCCTGCAAAGA	56	Alvarez-Iglesias et al. (2007)
5417	G→A	CCTCAATCACACTACTCCCCATA GGGGTGGGTTTTGTATGTTT	72	Alvarez-Iglesias et al. (2007)
8281- 8289del	deleção 9pb	TAGGGCCCGTATTTACCCTAT AAGAGGTGTTGGTTCTCTTAATCTTT	109	Alvarez-Iglesias et al. (2007)
8414	C→T	TACTACCGTATGGCCACCA GGGAGGTAGGTGGTAGTTTGTG	96	Alvarez-Iglesias et al. (2007)
9824	T→C	TTTTGTAGCCACAGGCTTCC GAGGAAAGTTGAGCCAATAATGA	53	Alvarez-Iglesias et al. (2007)
13759	G→A	GGAAGCCTATTGCGAGGATT GTTGTTTGAAGGGGGATG	64	Alvarez-Iglesias et al. (2007)
12705	T→C	CCCAAACATTAATCAGTTCTTCAA TCTCAGCCGATGAACAGTTG	102	(vanOven, Vermeulen, & Kayser, 2011)

Com o kit SsoFast™ EvaGreen® Supermix, a amplificação do marcador 8281-8289del foi realizada através da utilização do corante saturado EvaGreen®, que é um corante com fluorescência verde com espectros de emissão e excitação semelhantes ao do corante SYBR® Green I (possibilita a sua detecção em qualquer equipamento de PCR em Tempo Real sem necessidade de ajustes). A reação de amplificação foi efetuada para um volume final de 20 µl, contendo SsoFast EvaGreen supermix 1x, primer forward e reverse a 0,3 µM e aproximadamente 5 ng de DNA.

Com o kit AccuMelt™ HRM SuperMix o corante saturado utilizado foi o SYTO®9, que também é um corante com fluorescência verde. A reação de amplificação continha 1x AccuMelt HRM SuperMix (2x), 0,3 µM de *primer* forward e reverse e aproximadamente 5 ng de DNA, para uma volume final de 20 µl.

Com o kit Type-it® HRM™ PCR, o corante utilizado foi a EvaGreen®. A reação de amplificação continha 2x HRM PCR Master Mix a 1x, 0,3 µM de *primer* forward e reverse, e aproximadamente 5 ng de DNA, para um volume final de 25 µl.

As condições de amplificação utilizadas para a amplificação do DNA com os três kits, foram as recomendadas no manual do respetivo kit, como se encontram descritas no Anexo IV.

4. RESULTADOS E DISCUSSÃO

Neste capítulo irão ser descritos os resultados obtidos para as 38 amostras usando as metodologias: Minissequenciação, através do kit SNaPshot™ *Multiplex* e PCR em Tempo Real, recorrendo à análise HRM. Os resultados obtidos irão ser posteriormente comparados.

4.1. Minissequenciação

4.1.1. SNaPshot™ *Multiplex* kit – Seleção do polímero

O manual do SNaPshot™ (Biosystems, 2000) recomenda a utilização do polímero POP-4™ para o sequenciador automático ABI Prism® 310 Genetic Analyzer. No entanto, ao analisar as amostras com este polímero, verificou-se que os marcadores 3547 e 1736 do *multiplex* I (Figura 14 - A) e os marcadores 4820 e 4883 do *Multiplex* II (Figura 14 - B) sobreponham-se, ou seja, não era possível separar os alelos destes polimorfismos.

Os polímeros POP-4 e POP-6 apresentam concentrações de *uncross-linked dimethyl polyacrylamide* de 4% e 6%, respetivamente. Esta diferença de concentrações resulta num meio de separação menos viscoso para o polímero POP-4 do que para o polímero POP-6. Como os marcadores 3547/1736 e 4820/4883 apresentam tamanhos muito próximos, 19 pb e 22 pb, o polímero POP-6 é o mais indicado para separar os marcadores, uma vez que este é mais viscoso que o polímero POP-4, logo separa melhor os fragmentos (Figura 14 – C e D).

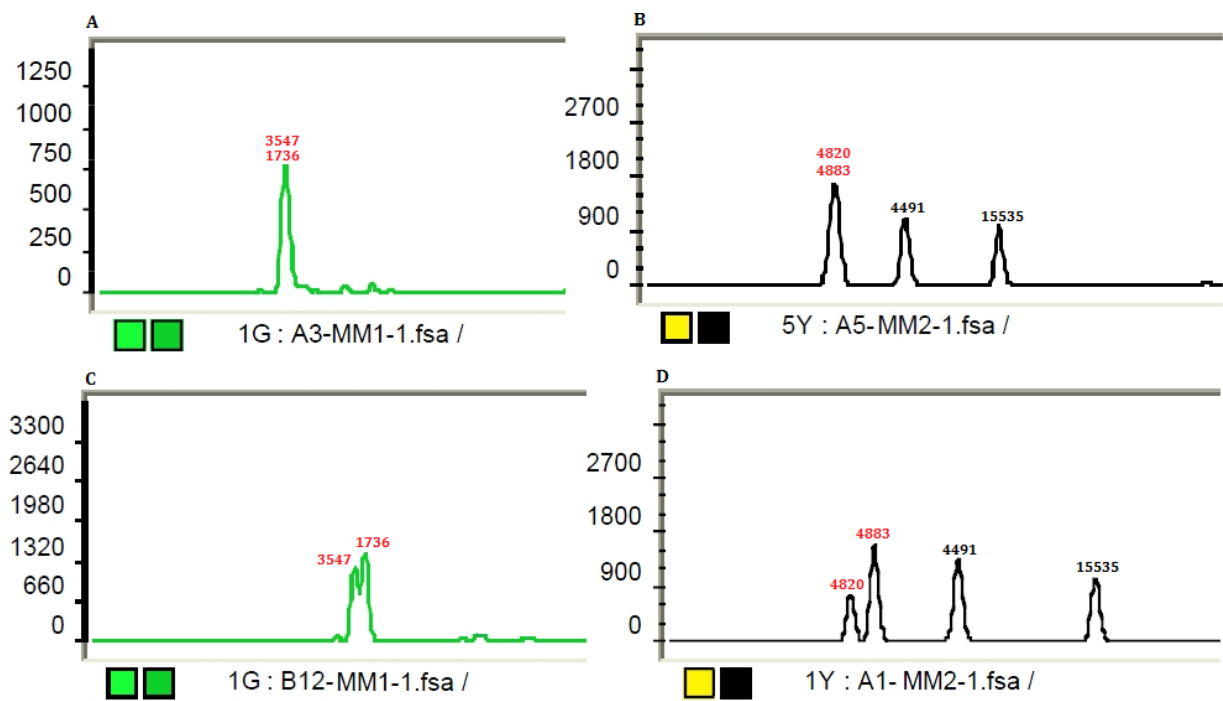


Figura 14 – Eletroferogramas dos marcadores 3547/1736 do *Multiplex I* e 4820/4883 do *Multiplex II* com o polímero POP-4 (A e B) e POP-6 (C e D).

4.1.2. Resultados

Para esta metodologia foram caracterizados 32 SNPs em 38 amostras, no sequenciador automático ABI Prism® 310 Genetic Analyzer utilizando o POP-6™ como polímero. Os resultados obtidos encontram-se resumidos na Tabela 5.

Tabela 5 – Tabela com os resultados da Minissequenciação da região codificante e sequenciação das regiões HVS I e HVS II.

Amostra ID	HVS I (início em 16000)	HVS II (mutações 73 e 263 estão presentes em todas as amostras)	1	1	1	2	3	3	3	4	4	4	4	5	8	9	1	1	1	1	1	1	1	1	1	1	1	1	1	9	HG									
			1	7	7	0	5	5	9	7	7	8	8	8	4	4	1	4	2	3	0	4	9	9	0	3	3	7	7	7		3	4	5	5	4	5	5	3	5
			T	G	A	C	A	T	C	C	G	A	C	G	G	C	T	G	G	C	A	G	G	T	A	T	T	G	G	T		G	A	C						
1	189, 217, 261	146	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	--	--	--	--	--	--	--	D	B4								
2	129, 144, 148, 222, 241, 265, 311, 343		--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								
5	129, 144, 148, 222, 241, 265, 294, 311, 343	89, 146, 208	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								
8	223, 295, 362	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
11	189	146	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	D	B4'5								
12	189, 223, 295	199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
13	148, 189, 223	152del, 185, 189	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	--									
15	140, 189, 217, 274	146, 150	C	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	D	B4c								
16	223, 291, 362, 390		--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
18	223, 291, 362, 390	146, 199	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
19	140, 189	103, 152, 204	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	G	C	--	--	--	--	--	--	D										
20	129, 172, 304, 311	249del	--	--	--	--	--	--	T	--	--	--	--	--	--	--	A	A	--	--	--	--	--	C	--	--	--	--	--	--	--	F								
21	069, 223, 261, 362, 390	320	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
22	189	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
23	189, 217, 247, 261, 274	146, 195	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	D	B4								
25	129, 144, 148, 157, 223, 241, 265, 278, 311, 343	89, 146, 195, 285	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								
26	189	146	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	D	B4'5								
27	086, 189, 217, 247, 261	146, 204, 207	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	D	B4								
28	157, 256, 304, 335	249del	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	--	R9c								
29	179, 189, 217, 261	146	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	D	B4								
30	129, 172, 304, 311	249del	--	--	--	--	--	--	T	--	--	--	--	--	--	--	A	A	--	--	--	--	--	C	--	--	--	--	--	--	--	F								
31	189, 223, 295	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
45	223, 362	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
46	129, 144, 148, 157, 223, 241, 265, 278, 311, 343	89, 146, 195, 285	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								
47	129, 144, 148, 223, 241, 265, 274, 311, 343, 362	89, 146, 173, 238	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								
48	223, 295, 362	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
49	157, 189, 256, 304, 342	249del	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	--	R9c								
53	126, 231, 311	234	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--									
54	093, 223, 295, 362	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
55	223, 291, 362	125, 127, 153	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
56	223, 261, 362, 390	152	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
57	223, 295, 362	146, 199	--	--	--	--	--	--	--	T	--	--	--	--	--	C	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M7b-g								
58	223, 291, 362, 390	125, 127, 235, 257	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
60	129, 172, 294, 304, 362	152, 249del	--	A	--	--	--	--	T	--	--	--	--	--	--	--	A	A	--	--	--	--	--	C	--	--	--	--	--	--	D									
62	157, 256, 304, 335	249del	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	A	--	--	--	--	--	C	--	--	--	--	--	--	--	R9c								
72	129, 144, 148, 222, 241, 265, 311, 343		--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								
78	189, 223, 291	195	--	--	--	--	--	--	--	--	--	--	A	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M9								
88	129, 144, 148, 222, 241, 265, 311, 343	89, 92, 93, 146, 208	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--	--	--	--	--	--	--	--	--	--	M*								

Dos resultados obtidos da minissequenciação da região codificante e sequenciação das regiões HVS I e HVS II (resultados obtidos no âmbito do projeto da FCT: POCI/ANT/6178/2004) foi possível determinar os haplogrupos (HG) das amostras de acordo com a árvore filogenética construída para os 32 SNPs em estudo (Figura 13).

Nas amostras 19 e 53 não foi possível determinar o haplogrupo. Isto deve-se ao facto destas não conterem os polimorfismos 10398 e 10400 que definem os macrohaplogrupos N e M, respetivamente, que dão origem às linhagens Europeias e Asiáticas (VanOven & Kayser, 2009). A ausência destes polimorfismos (assinalados a amarelo na Tabela 5) pode indicar que as amostras possuem descendência Africana. Contudo, existem outros polimorfismos nas amostras que sugerem o contrário, ou seja, colocam as amostras em linhagens Asiáticas e Europeias.

A amostra 19 apresenta o polimorfismo 12705 que é específico do haplogrupo R (haplogrupo mais comum da Euro-Ásia ocidental) e o polimorfismo 12358 que só aparece em haplogrupos europeus e asiáticos. Por sua vez, a amostra 53 apresenta o polimorfismo 5417 que é específico do haplogrupo N9 (característico da Ásia) e R12 (encontrado na Austrália). Estes polimorfismos presentes nas amostras poderão definir um novo haplogrupo. Deste modo, será necessário testar as amostras para outros marcadores mitocondriais e em último recurso, sequenciar todo o genoma mitocondrial.

Relativamente à amostra 60, também não foi possível determinar o haplogrupo. Os polimorfismos 10398, 12705, 3970, 10310 da região codificante e os polimorfismos 16304 e 249del das regiões HVS I e HVS II indicam que esta amostra pertence ao haplogrupo F. Mas nesta amostra também estão presentes dois polimorfismos, 1719 e a deleção de 9 pb (indicados a azul na Tabela 5), que não se encontram em nenhuma ramificação do haplogrupo F (VanOven & Kayser, 2009). Estes polimorfismos podem caracterizar um novo subhaplogrupo, pertencente ao haplogrupo F. Deste modo, será necessário testar outros marcadores mitocondriais e em último recurso, sequenciar todo o genoma mitocondrial.

Dos resultados da minissequenciação da amostra 13 verificou-se que esta amostra apresentava os polimorfismos 10400, que é específico do macrohaplogrupo M e 4883, polimorfismo que só se encontra nos haplogrupos M80'D e H1e3 (assinalados a vermelho na Tabela 5). A presença destes polimorfismos indica que a amostra pode pertencer ao haplogrupo M80'D. No entanto, os polimorfismos das regiões HVS I e HVS II não confirmam este haplogrupo. Como os marcadores estudados não são suficientes para classificar a amostra 13, não foi possível determinar o haplogrupo.

As amostras 28, 49 e 62 encontram-se classificadas na Tabela 5 como pertencentes ao haplogrupo R9c, mas os polimorfismos da região HVS I presentes nestas amostras levam a classificá-las como R9c1a. De forma a confirmar esta classificação, será necessário genotipar outros marcadores mitocondriais da região codificante presentes no haplogrupo R9c1a.

4.2. PCR em Tempo Real: Análise HRM

4.2.1. Análise comparativa com os três kits de amplificação

Ao testar os kits de amplificação: SsoFast™ EvaGreen® Supermix (Bio-Rad), AccuMelt™ HRM SuperMix (Quanta Biosciences™) e Type-it® HRM™ PCR (Qiagen) no CFX96™ Real-Time PCR (Bio-Rad), obtiveram-se os seguintes resultados:

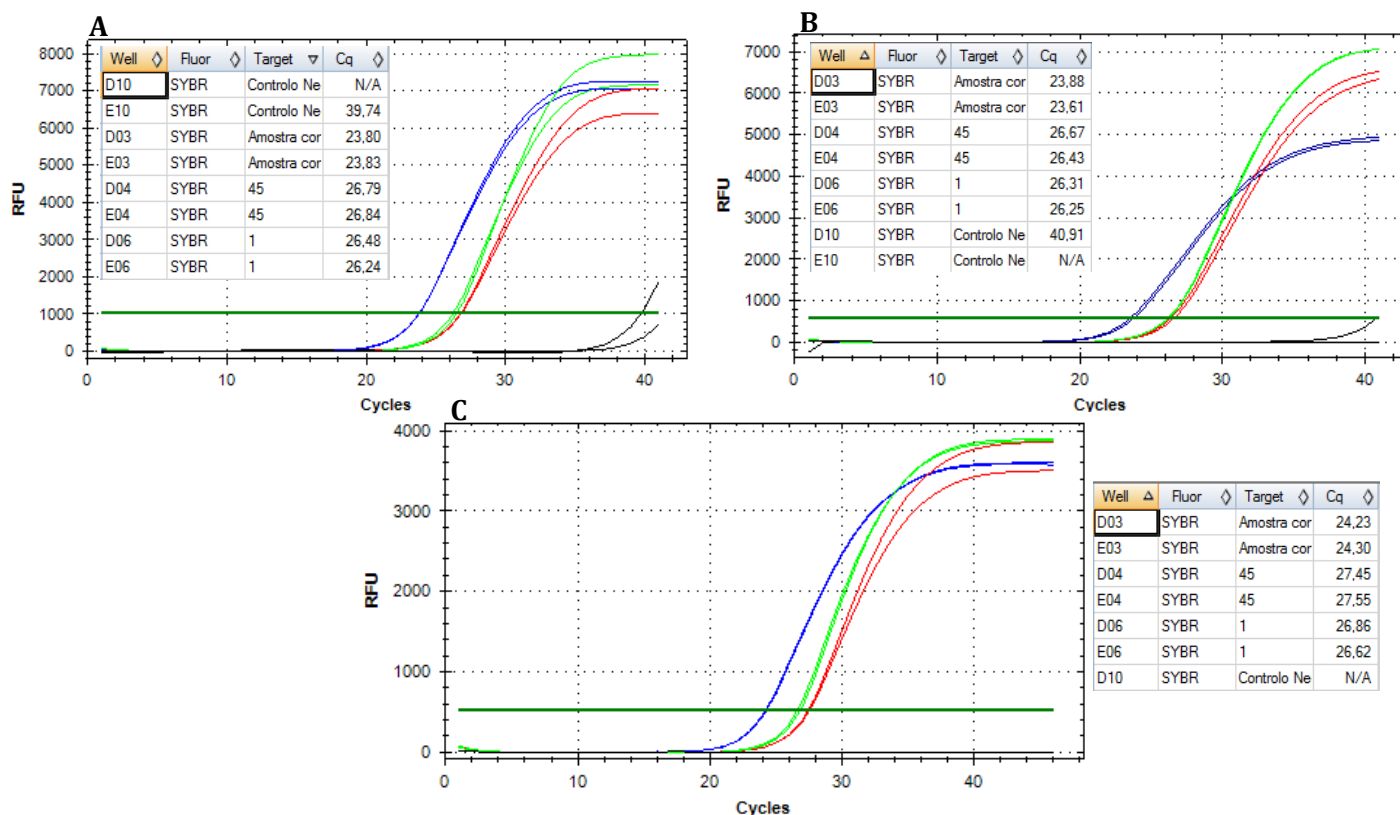


Figura 15 – Representação gráfica da curva de amplificação das amostras teste nos três kits. **A** – SsoFast™ EvaGreen® Supermix; **B** – AccuMelt™ HRM SuperMix; **C** – Type-it® HRM™ PCR. **Preto** – controlo negativo; **Azul** – amostra controlo; **Verde** – amostra 1; **Vermelho** – amostra 45.

Os resultados obtidos podem ser visualizados sob a forma gráfica (Figura 15), podendo-se observar no eixo dos x o número de ciclos de amplificação e no eixo dos y as unidades de fluorescência (RFU) da reação de amplificação, que é proporcional à quantidade de produto amplificado.

Ao visualizar os resultados obtidos, verificou-se que os kits de amplificação SsoFast™ EvaGreen® Supermix e AccuMelt™ HRM SuperMix eram os que apresentavam melhores resultados em relação ao kit Type-it® HRM™ PCR, uma vez que os níveis de fluorescência (RFU) eram mais elevados e os valores de C_T (número de ciclos onde se verifica um aumento significativo na quantidade de produtos amplificados) eram um pouco mais baixos. Deste modo, níveis elevados de fluorescência, conjugados com baixos valores de C_T representam duas vantagens.

Ao efetuar-se a comparação dos dois primeiros kits, verificou-se que estes proporcionavam valores de fluorescência e C_T similares, mas o kit EvaGreen® Supermix

apresentava curvas de amplificação com melhor perfil, ou seja, as fases exponencial e *plateau* encontravam-se bem definidas.

4.2.2. Resultados

As 38 amostras em estudo foram caracterizadas usando 9 SNPs, utilizando o kit EvaGreen® Supermix. Para a obtenção de resultados na análise HRM, foi utilizada uma amostra controlo em que os SNPs em estudo foram previamente estudados usando o SNaPshot®. Com a utilização da amostra controlo na análise HRM, é possível efetuar uma comparação das curvas de *melting* desta amostra, com as curvas de *melting* das amostras em estudo (Figura 16); ou seja, consoante a temperatura de *melting* que as amostras apresentem, estas agrupam-se formando *clusters* e, desta forma, será possível determinar se as amostras apresentam ou não o polimorfismo.

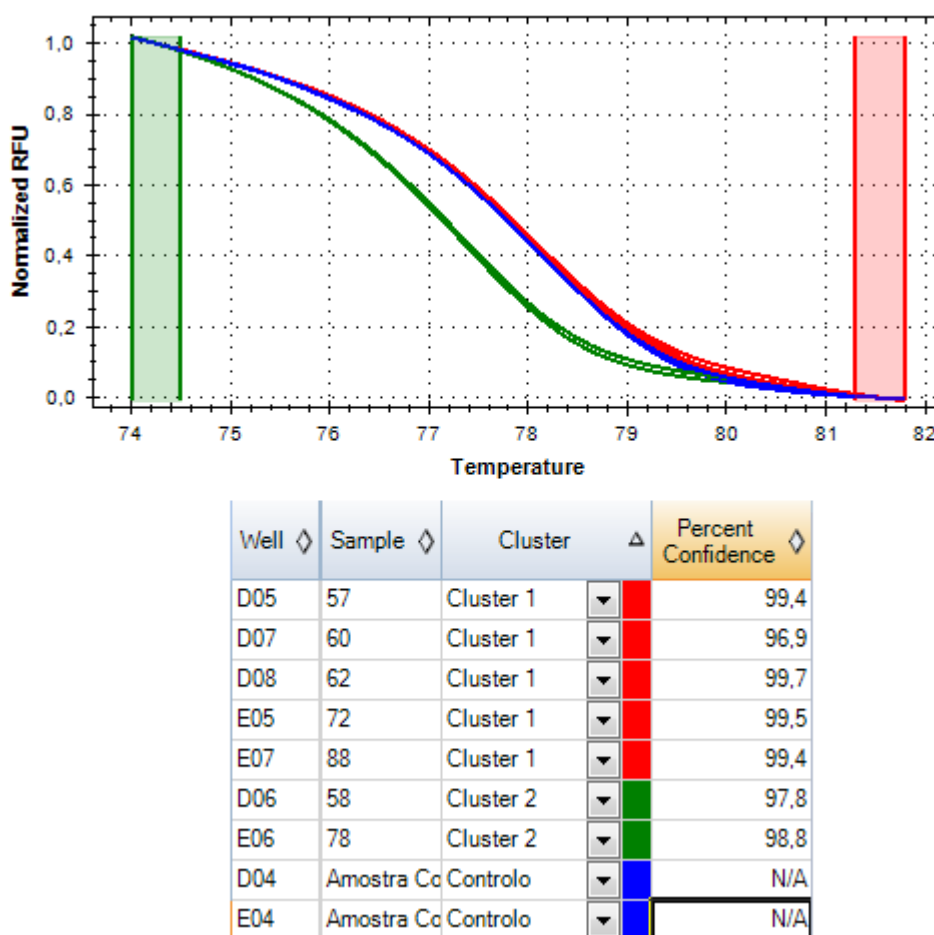


Figura 16 – Exemplo das curvas de *melting* normalizadas e sua respetiva tabela de identificação para o marcador 4491, obtidas através da análise HRM.

Os resultados obtidos para as 38 amostras em 9 marcadores moleculares encontram-se resumidos na Tabela 6. Nesta tabela, considera-se negativo o alelo ancestral e positivo o alelo mutado.

[illegible]

Como a determinação da Temperatura de *Annealing* (T_a) é um passo essencial para a obtenção de um bom resultado na análise HRM, foi utilizada a aplicação 'Ta Calculator' do software CFX Manager™ 2.0 (Bio-Rad) na determinação da T_a para os diferentes marcadores.

No marcador 2092, a temperatura calculada foi de 56°C. Ao genotipar este marcador em algumas amostras com a $T_a=56^\circ\text{C}$, verificou-se que a amplificação não era específica pois as curvas de *melting* (Figura 17 – A) apresentavam dois picos principais com temperaturas de *melting* (T_m) distintas. Na análise HRM (Figura 17 – B) também não se pode tirar qualquer conclusão em relação aos resultados obtidos, pois o perfil da curva de *melting* não deu segurança em relação aos resultados obtidos.

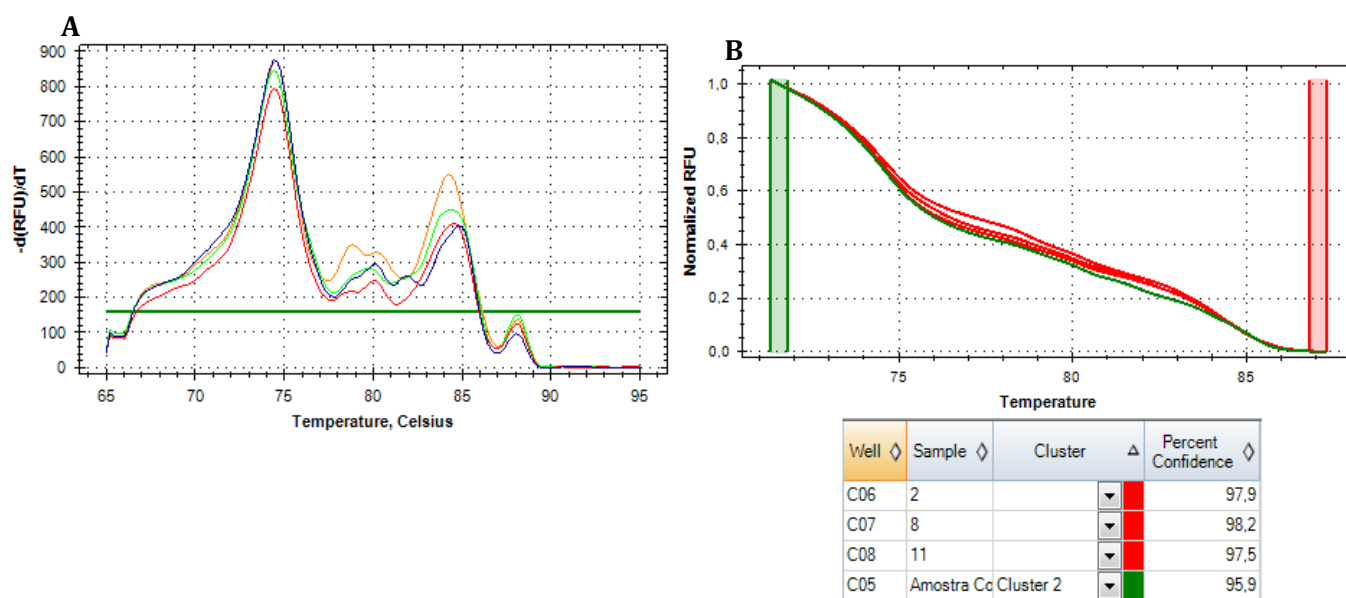


Figura 17 – Exemplo de resultados obtidos para o marcador 2092 com $T_a=56^\circ\text{C}$.

A – Curva de *melting*;

B – Curva de *melting* normalizada, na análise HRM e sua respectiva tabela de identificação.

De forma a tentar contornar esta situação realizou-se um gradiente de temperaturas para este marcador. Através da visualização das curvas de *melting* (Figura 18 – A), verificou-se que a melhor temperatura de *annealing* era a de 60°C , pois a curva de *melting* apresentava apenas um único pico.

Desta forma, conseguiu-se genotipar as 38 amostras para o marcador 2092.

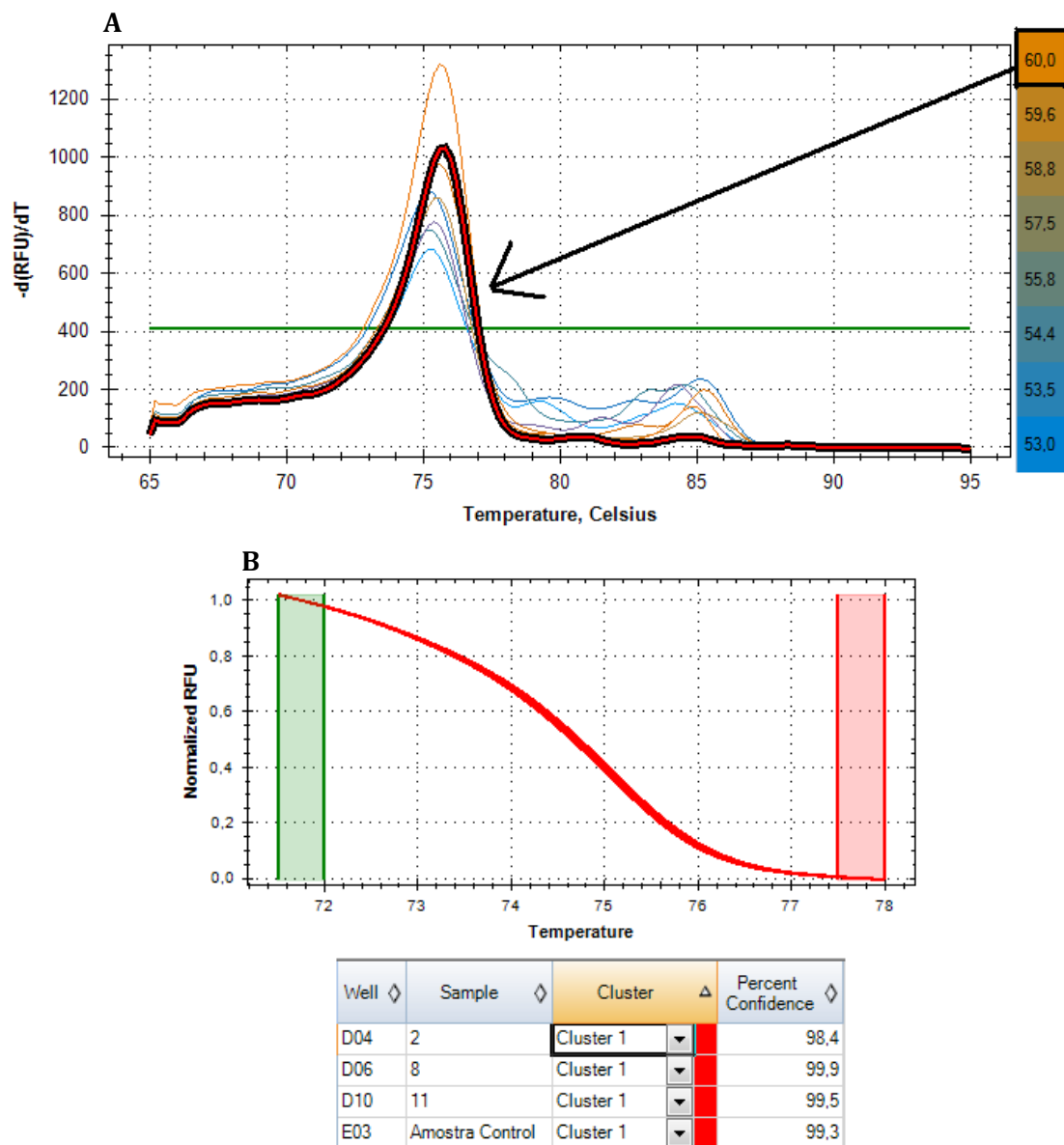


Figura 18 – Resultados obtidos após o gradiente de temperatura para o marcador 2092.

A – Curvas de *melting* do gradiente de temperatura.

B – Curva de *melting* normalizada, na análise HRM e sua respectiva tabela de identificação, com temperatura de *annealing* de 60°C.

4.3. Análise comparativa das metodologias

Ao comparar os resultados obtidos nas duas metodologias, verificou-se uma discordância de resultados. A Tabela 7 demonstra a discrepância dos resultados (indicados a vermelho) na amostra 13 para o marcador 1119, nas amostras 28, 49 e 62 para o marcador 8414 e na amostra 60 para o marcador 13759. Nas restantes amostras e marcadores estudados, os resultados correspondem nas duas metodologias.

Tabela 7 – Tabela comparativa dos resultados de 3 marcadores no SNaPshot™ e na análise HRM.
Legenda: - alelo ancestral; + alelo mutado.

Amostra ID	SNaPshot™			Análise HRM		
	1119	8414	13759	1119	8414	13759
Amostra controlo	-	-	-	-	-	-
13	-	-	-	+	-	-
28	-	-	-	-	+	-
49	-	-	-	-	+	-
60	-	-	-	-	-	+
62	-	-	-	-	+	-

No marcador 8414, as amostras 28, 49 e 62 deram resultado positivo na análise HRM e negativo no SNaPshot™. Esta diferença de resultados pode ser explicada pela não especificidade dos *primers* na análise HRM para detetar o polimorfismo 8414.

Ao analisar as amostras 28, 49 e 62 para o marcador 8414 no *software* HRM, observa-se uma separação evidente de grupos (*cluster*) entre as curvas de *melting* das amostras em estudo e a curva de *melting* da amostra controlo (Figura 19). Esta diferença nas curvas de *melting* não se deve à presença do polimorfismo 8414, pois este polimorfismo é específico para o haplogrupo D4 e as amostras foram classificadas como haplogrupo R9c através da minissequenciação e sequenciação das regiões HVS I e HVS II (Tabela 5). Assim, a diferença nas curvas de *melting* pode dever-se à existência de outro polimorfismo nos *amplicons* e não do polimorfismo 8414.

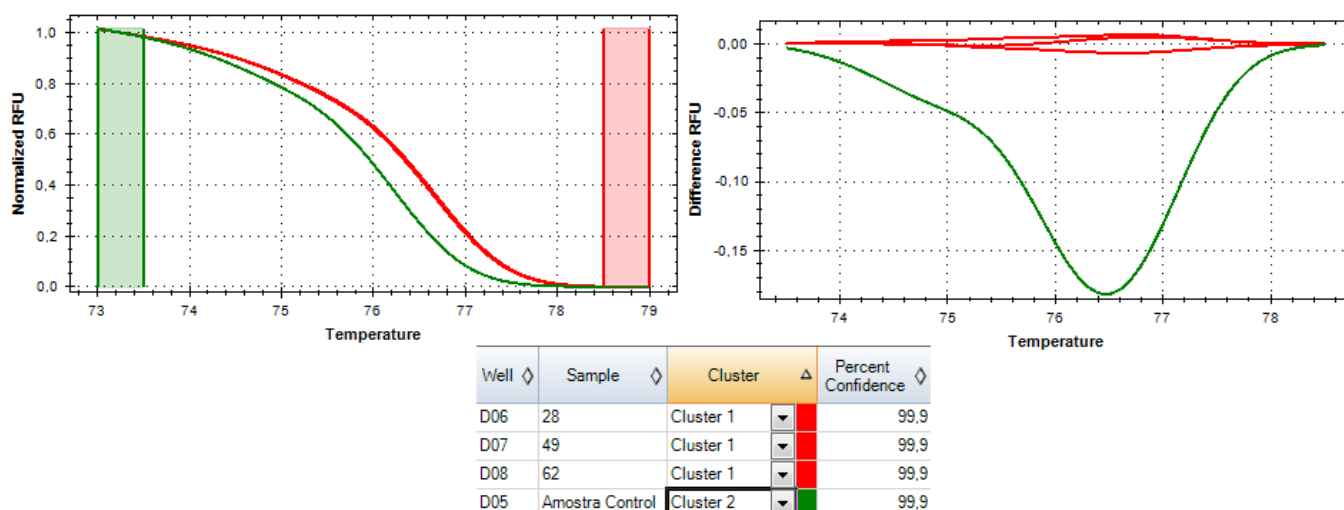


Figura 19 – Curva de *melting* normalizada, gráfico de diferenças obtido da curva de *melting* normalizada e tabela de identificação para as amostras 28, 49 e 62.

Ao verificar qual o segmento da região codificante do DNAm_t gerado pelos *primers* utilizados para detetar o polimorfismo 8414, através da comparação dos *primers* com a sequência de referência revista (rCRS), verificou-se que o polimorfismo 8440, presente no haplogrupo R9c1a, encontrava-se no *amplicon* (Figura 20). Deste modo, o polimorfismo detetado pela análise HRM será o polimorfismo 8440 e não o 8414, pois as amostras pertencem ao haplogrupo R9c e com os polimorfismos presentes na região HVS I é possível classificá-las como R9c1a.

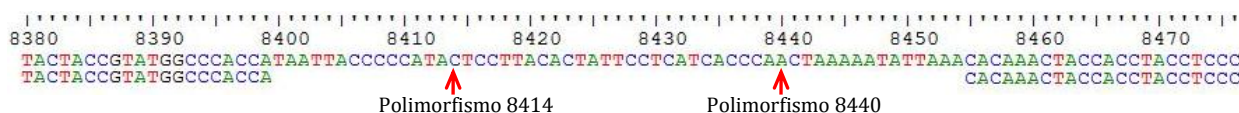


Figura 20 – Comparação da sequência de referência com os *primers* utilizadas para detetar o polimorfismo 8414.

Como a análise HRM se baseia na comparação de curvas de *melting*, não determinando qual o polimorfismo presente em determinada posição, é fundamental a utilização de *primers* que amplifiquem apenas um SNP. Por vezes, dois ou mais SNPs podem estar localizados muito próximos entre si tornando-se difícil desenhar um *primer* para localizar um único SNP. Contudo é possível amplificar uma região que contenha múltiplos SNPs e realizar a análise HRM, desde que todos os controlos sejam incluídos.

Em relação à discrepância de resultados na amostra 13 para o marcador 1119 e na amostra 60 para o marcador 13759, também pode ser explicada pela não especificidade dos *primers* utilizados para a análise HRM. Como não foi possível determinar o haplogrupo destas amostras, o polimorfismo presente nos *amplicons* não foi identificado.

4.4. Distribuição de haplogrupos

A identificação dos haplogrupos e subhaplogrupos foi efetuada com base nos resultados obtidos a partir do estudo de 32 SNPs da região codificante do genoma mitocondrial e apoiada pela análise de sequenciação das regiões hipervariáveis (DNA não codificante) HVS I e HVS II previamente realizada no âmbito do projeto da FCT: POCI/ANT/6178/2004 (Tabela 5).

A análise dos polimorfismos da região codificante e não codificante do DNA mitocondrial permitiu a obtenção de resultados que podem constituir uma aproximação a um estudo populacional, embora o número de amostras seja baixo. Nas 38 amostras estudadas observou-se um total de 7 haplogrupos, definidos de acordo com VanOven and Kayser (2009), em que 2 haplogrupos pertencem ao macrohaplogrupo M e 5 haplogrupos pertencem ao macrohaplogrupo N.

Das 22 amostras caracterizadas como macrohaplogrupo M, 8 amostras pertencem ao subhaplogrupo M7b-g, sendo este característico da Ásia oriental; 7 amostras pertencem ao haplogrupo M9, que também é característico da Ásia oriental e central. As restantes 7 amostras foram classificadas como M*, ou seja, estas amostras pertencem a um subhaplogrupo que não foi possível identificar, devido ao reduzido número de marcadores estudados.

Para as 12 amostras classificadas como macrohaplogrupo N, 7 amostras pertencem ao subhaplogrupo B4'5, das quais 4 amostras encontram-se classificadas como subhaplogrupo B4 e 1 amostra como subhaplogrupo B4c. Os subhaplogrupos B4 e B5 encontram-se espalhados pela Ásia ocidental e no sudoeste Asiático. Nas restantes 5 amostras, 2 pertencem ao haplogrupo F que é bastante comum na Ásia ocidental e no sudoeste Asiático e 3 ao subhaplogrupo R9c que se encontra por todo o arquipélago malaio.

Um estudo anteriormente realizado a Este da Indonésia (Mona et al., 2009) demonstra a frequência dos haplogrupos mitocondriais presentes em Timor-Leste e regiões vizinhas. Este estudo baseou-se na utilização da região HVS I e a deleção de 9 pb do DNA mitocondrial para determinar os haplogrupos presentes em 330 amostras provenientes da Indonésia oriental. Das 330 amostras utilizadas, 38 provinham de Timor-Leste, tendo sido identificados os haplogrupos e subhaplogrupos B, B4a, B5b, F1a, R9c, P1, Q1, E1a e B4a1a1a. Os haplogrupos Q1 e E1a (o subhaplogrupo E1a pertence ao haplogrupo M9) são os que apresentavam maior frequência no conjunto de amostras estudadas.

Ao comparar os haplogrupos obtidos por Mona e colaboradores (2009) com os haplogrupos obtidos neste trabalho para as amostras de Timor-Leste, verificou-se que existe uma concordância para os haplogrupos B, R9c, E e F. No entanto, existe uma diferença entre os dois estudos que importa salientar, que é a ausência do haplogrupo Q no presente estudo, uma vez que este é descrito por Mona e seus colaboradores (2009) como sendo um dos haplogrupos mais frequente. Esta ausência será apenas devida ao acaso, resultante do reduzido número de amostras utilizado. Deste modo, o aumento da amostragem, assim como do número de SNPs a analisar, permitirá uma melhor caracterização desta população.

5. CONCLUSÕES

A identificação genética de material biológico é rotineiramente realizado através da análise de STRs do DNA nuclear. Contudo, em casos em que o DNA se encontra degradado ou em quantidades exíguas, a análise do DNAm é muitas vezes utilizada devido ao seu elevado número de cópias por célula.

Nos últimos anos o interesse nos SNPs aumentou significativamente, devido à sua abundância no genoma humano e, simultaneamente, às novas tecnologias de detecção de SNPs que têm vindo a ser desenvolvidas para localizar e caracterizar variações em locais específicos do genoma humano. A escolha do método apropriado para genotipagem de SNPs na área forense é crítica para o sucesso de um estudo, devendo ter em conta diversos aspetos, tais como a sensibilidade, precisão e reprodutibilidade.

Neste trabalho foi estudada uma amostra populacional de 38 indivíduos naturais de Timor-Leste, recorrendo a marcadores bi-alélicos localizados na região codificante do DNA mitocondrial, de forma a comparar duas metodologias de genotipagem de SNPs (Minissequenciação, com a utilização do kit SNaPshot™ *multiplex* e PCR em Tempo Real, através da análise HRM) e determinar os haplogrupos e subhaplogrupos presentes na amostragem em estudo, tendo também em consideração a sequenciação das regiões HVS I e HVS II do DNA mitocondrial.

Com este estudo, conseguiu-se obter algumas conclusões, descritas:

Da comparação das duas metodologias de genotipagem de SNPs utilizadas neste estudo, verificou-se que a minissequenciação se revela uma metodologia mais precisa, robusta, fiável e com possibilidade de análise de vários *loci* em simultâneo – *multiplex*. Estas vantagens justificam uma maior utilização deste método nos laboratórios de Genética Forense. A estas vantagens acresce o facto do método de detecção utilizado (instrumentos de eletroforese capilar) ser o correntemente usado na análise dos marcadores de eleição em Genética Forense (STRs). Contudo, este é um método com um procedimento mais complexo, o que resulta num tempo de resposta mais elevado.

Em face das vantagens e desvantagens da minissequenciação, considera-se vantajosa a utilização da PCR em Tempo Real, seguida de análise HRM, como metodologia a ser integrada na rotina laboratorial como método de pré-seleção. No entanto, deverá ter-se

em consideração a utilização de amostras controlos para cada um dos polimorfismos a estudar. Esta opção do uso prévio desta metodologia prende-se com as vantagens desta, designadamente, ser um método simples, rápido, pouco dispendioso e de rápida utilização e o procedimento ser efetuado num único passo.

As desvantagens da análise HRM, designadamente o facto de não fornecer informação sobre a mutação específica presente no *amplicon* e não ser facilmente aplicada em *multiplex*, faz com que esta metodologia não possa ser facilmente aplicada na rotina laboratorial da Genética Forense da mesma forma que o SNaPshot™.

A utilização desta técnica em Genética Forense poderá tornar a genotipagem de SNPs um método menos dispendioso, uma vez que o equipamento utilizado já se encontra disponível em quase todos os laboratórios forenses. A única limitação prende-se com a necessidade de adquirir o *software* de *Precision Melt Analysis*™ utilizado para efetuar a análise HRM.

Relativamente à distribuição de haplogrupos da amostra da população de Timor-Leste, verificou-se que o número de marcadores utilizados para a sua caracterização demonstrou não ser suficiente, uma vez que 7 amostras das 38 estudadas ficaram classificadas como haplogrupo M*, ou seja, não foi possível definir o subhaplogrupo. Desta forma, é necessário uma maior resolução filogenética usando um maior número de marcadores, de modo a melhor caracterizar as amostras em subhaplogrupos mais específicos. A caracterização dos SNPs do DNA mitocondrial deve recair não só na discriminação de linhagens *major*, mas também, na utilização de um mínimo de polimorfismos que permitam um aumento significativo do poder de discriminação.

Por isso, considera-se que a determinação dos haplogrupos permitiu a obtenção de resultados que podem constituir elementos importantes para um futuro estudo populacional.

Em suma, acredita-se que qualquer uma das metodologias estudadas é uma mais-valia para a Genética Forense e que têm contribuído para o avanço científico na área; sendo que o trabalho explorado ao longo desta dissertação indica que a minissequenciação, usado o kit *SNaPshot*™ *multiplex*, é o método mais vantajoso para estudos e investigações forenses.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Alvarez-Iglesias, V., Jaime, J. C., Carracedo, A., & Salas, A. (2007). Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet*, 1(1), 44-55.
- Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., . . . Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806), 457-465.
- Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*, 23(2), 147.
- Bio-Rad Laboratories, I. (2006). Real-Time PCR Applications Guide.
- Biosystems, A. (2000). ABI PRISM® SNaPshot™ Multiplex Kit - Protocol.
- Biosystems, A. (2001). ABI PRISM® SNaPshot™ Multiplex System.
- Branco, J. D. A. d. C. (1987). A Ordem de São Domingos e as origens de Timor. *Sociedade Histórica da Independência de Portugal*.
- Brandstätter, A., Parsons, T., & Parson, W. (2003). Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med*, 291-298.
- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177-186.
- Butler, J. M. (2009). *Fundamentals of Forensic DNA Typing*: Academic Press.
- Butler, J. M. (2012). *Advanced Topics in Forensic DNA Typing: Methodology*: Academic Press.
- Cann, R., Stoneking, M., & Wilson, A. (1987). Mitochondrial DNA and human evolution. *Nature*, 325, 31-36.
- Carracedo, A. (2005). *Forensic DNA typing protocols*: Human Press Inc.
- Centeno, R., & Novais, R. (2006). *Timor-Leste: da Nação ao Estado*: Edições Afrontamento.

- Dobrowolski, S. F., Gray, J., Miller, T., & Sears, M. (2009). Identifying sequence variants in the human mitochondrial genome using high-resolution melt (HRM) profiling. *Human Mutation*, 30(6), 891-898.
- Dorak, M. T. (2007). *Real-time PCR*: Taylor & Francis Group.
- Erali, M., Voelkerding, K., & Wittwer, C. (2008). High Resolution Melting applications for clinical laboratory medicine. *Exp Mol Pathol*, 85(1), 50-58.
- Farrar, J., Reed, G., & Wittwer, C. (2010). *High-resolution melting curve analysis for molecular diagnostics*: Academic Press.
- Fox, J., & Soares, D. (2003). *Out of the Ashes: Deconstruction and Reconstruction of East Timor*: Australian National University.
- Gidlöf, O., Burvall, S., Edvinsson, L., Montelius, M., Allen, M., & Molin, M. (2009). Complete discrimination of six individuals based on high-resolution melting of hypervariable regions I and II of the mitochondrial genome. *BioTechniques*, 47, 671-678
- Goodwin, W., Linacre, A., & Hadi, S. (2011). *An Introduction to Forensic Genetics* (Second ed.): Wiley
- Gundry, C., Vandersteen, J., Reed, G., Pryor, R., Chen, J., & Wittwer, C. (2003). Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clinical Chemistry*, 49(3), 396-406.
- Hattori, R., Gomes, M., Ajo, F., & Belo, N. (2005). *The Ethnolinguistic Situation in East Timor*. Paper presented at the International Graduate Student Conference Series, University of Hawaii.
- Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. (1996). Real time quantitative PCR. *Genome Research*, 986-994.
- Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR Analysis: Real-time Monitoring of DNA Amplification Reactions. *Biotechnology Nature Publishing Company*, 11, 1026-1030.
- Hull, G. (2004). The Languages of East Timor: Some Basic Facts Acedido a 27 October, 2011, de <http://www.ramelau.com/tetum/facts.php>
- Human_Genome_Project_Information. (2010). Genetic Anthropology, Ancestry, and Ancient Human Migration Acedido a 23 January, 2012, de http://www.ornl.gov/sci/techresources/Human_Genome/elsi/humanmigration.shtml
- Javed, R., & Mukesh. (2010). Current research status, databases and application of single nucleotide polymorphism. *Pak J Biol Sci*, 13(13), 657-663.
- Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature*, 316(6023), 76-79.

- Kayser, M. (2007). Uni-parental markers in human identity testing including forensic DNA analysis. *BioTechniques*, 43, xvi-xxi.
- Kwok, P.-Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet*, 2, 235-258.
- Mackay, I. M. (2007). *Real-Time PCR in Microbiology - From Diagnosis to Characterization* Caister Academic Press
- Mannis vanOven, & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*, 30(2), E386-E394.
- Martino, A., Mancuso, T., & Rossi, A. M. (2010). Application of High-Resolution Melting to Large-Scale, High-Throughput SNP Genotyping: A Comparison with the TaqMan® Method. *Journal of Biomolecular Screening*.
- McNevin, D., Bate, A., Daniel, R., & Walsh, S. J. (2011). A preliminary mitochondrial DNA SNP genotyping assay for inferring genealogy. *Australian Journal of Forensic Sciences*, 43, 39-51.
- Mona, S., Grunz, K. E., Brauer, S., Pakendorf, B., Castri, L., Sudoyo, H., . . . Kayser, M. (2009). Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Molecular Biology and Evolution*, 1865-1877.
- Mosquera-Miguel, A., Álvarez-Iglesias, V., Cerezo, M., Lareu, M. V., Carracedo, Á., & Salas, A. (2009). Testing the performance of mtSNP minisequencing in forensic samples. *Forensic science international. Genetics*, 3(4), 261-264.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, 51 Pt 1, 263-273.
- National_Geographic. (1996). The Genographic Project Acedido a 24 October, 2011, de <https://genographic.nationalgeographic.com/genographic/lan/en/signposts.html>
- O'Connor, S., Spriggs, M., & Veth, P. (2002). Excavation at Lene Hara Cave establishes occupation in East Timor at least 30,000–35,000 years ago. *Antiquity Publications*, 76, 45-50.
- Parsons, T. J., & Coble, M. D. (2001). Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croatian Medical Journal*, 42(3), 304-309.
- Pinheiro, M. (2010). *Genética Forense: Perspectivas da Identificação Genética*: Universidade Fernando Pessoa.
- Quintáns, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V., & Carracedo, A. (2004). Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int*, 140, 251-257.

- Reed, G., Kent, J., & Witwer, C. (2007). High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics*, 8(6), 597-607.
- Richards, M., Bandelt, H.-J., Kivisild, T., & Oppenheimer, S. (2006). A Model for the Dispersal of Modern Humans out of Africa. In H.-J. Bandelt, V. Macaulay & M. Richards (Eds.), *Human Mitochondrial DNA and the Evolution Homo sapiens* (Vol. 18, pp. 227-268): Springer Berlin Heidelberg.
- Salas, A., Bandelt, H. J., Macaulay, V., & Richards, M. B. (2007). Phylogeographic investigations: the role of trees in forensic genetics. *Forensic Sci Int*, 168(1), 1-13.
- SEKM. (2011). Governo de Timor-Leste Acedido a 27 September, 2011, de <http://timor-leste.gov.tl/?>
- Sobrino, B., Brión, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic science international*, 154(2-3), 181-194.
- Taylor, S., Scoot, R., Kurtz, R., Fisher, C., Patel, V., & Bizouram, F. (2010). A practical guide to High Resolution Melt Analysis Genotyping. In I. Bio-Rad Laboratories (Ed.).
- Thompson, T., & Black, S. (2007). *Forensic Human Identification: An Introduction*: CRC Press.
- Torrioni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., . . . Wallace, D. C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*, 53(3), 563-590.
- Torrioni, A., Schurr, T. G., Yang, C. C., Szathmary, E. J., Williams, R. C., Schanfield, M. S., . . . et al. (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics*, 130(1), 153-162.
- Tully, G., Sullivan, K., Nixon, P., Stones, R., & Gill, P. (1996). Rapid Detection of Mitochondrial Sequence Polymorphisms Using Multiplex Solid-Phase Fluorescent Minisequencing. *Genomics*, 34(1), 107-113.
- Twyman, R. (2003). Techniques patents for SNP genotyping. *Pharmacogenomics*, 4(1), 67-79.
- Twyman, R. (2005). Single Nucleotide Polymorphism (SNP) Genotyping Techniques—An Overview. *Encyclopedia of Diagnostic Genomics and Proteomics*.
- Umetsu, K., Tanaka, M., Yuasa, I., Adachi, N., Miyoshi, A., Kashimura, S., . . . Osawa, M. (2005). Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations. *Electrophoresis*, 26(1), 91-98. doi: 10.1002/elps.200406129
- UNDP. (2002). Ukun Raski A'an - The way ahead (pp. 70-75).
- VanOven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2), E386-E394.

- vanOven, M., Vermeulen, M., & Kayser, M. (2011). Multiplex genotyping system for efficient inference of matrilineal genetic ancestry with continental resolution. *Investig Genet*, 2, 6.
- Vidiani. (2011). Maps of all countries in one place Acedido a 05 September, 2011, de <http://www.vidiani.com/?p=3394>
- Vossen, R., Aten, E., Roos, A., & Dunnen, J. (2009). High-resolution melting analysis (HRMA): more than just sequence variant screening. *Human Mutation*, 30(6), 860-866.
- Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex-100 as a Medium for Simple Extraction of DNA for Pcr-Based Typing from Forensic Material. *Biotechniques*, 10(4), 506-513.
- Watson, E., Forster, P., Richards, M., & Bandelt, H. J. (1997). Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet*, 61(3), 691-704. doi: 10.1086/515503
- Wittwer, C. (2009). High-resolution DNA melting analysis: advancements and limitations. *Human Mutation*, 30(6), 857-859.
- Wittwer, C., Herrmann, M., Moss, A., & Rasmussen, R. (1997). Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, 22, 130-138.
- Wittwer, C., Reed, G., Gundry, C., Vandersteen, J., & Pryor, R. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry* 49, 853-860
- Wittwer, C., Ririe, K., Andrew, R., David, D., Gundry, R., & Balis, U. (1997). The LightCycler™: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques*, 22(1), 176-181.
- Zuccarelli, G., Alechine, E., Caputo, M., Bobillo, C., Corach, D., & Sala, A. (2010). Rapid screening for Native American mitochondrial and Y-chromosome haplogroups detection in routine DNA analysis. *Forensic Sci Int Genet*.

7. ANEXOS

Anexo I – Extração de DNA pelo método adaptado de *Chelex*®100

O procedimento de extração de DNA recorrendo ao método adaptado de *Chelex*® 100 compreende os seguintes passos:

1. A tubos *ependorf* de 1,5 ml (estéreis) adiciona-se 10 µl de sangue total;
2. Para cada *ependorf* pipeta-se 1 ml de diH₂O (água desionizada) e agita-se suavemente;
3. Incuba-se à temperatura ambiente durante 20 minutos;
4. Agita-se no *Vórtex* durante 5-10 segundos, cuidadosamente;
5. Centrifuga-se a 14000 rpm (rotações por minuto) durante 3 minutos;
6. Retira-se o sobrenadante, deixando 25 µl no tubo *ependof*;
7. Adiciona-se 150 µl de solução de *Chelex* 5%;

Nota: Para este passo é necessário cortar a extremidade das pontas das pipetas para ficarem mais largas; a solução de *Chelex* deverá estar no agitador magnético.

8. Adiciona-se 10 µl de *Proteinase K*;
9. Incuba-se durante 30 minutos a 56 °C no termobloco
10. Agita-se vigorosamente no *Vórtex* 5-10 segundos;
11. Incuba-se num banho em ebulição (90-99 °C) durante 8 minutos. **Nota:** com uma agulha estéril, perfuraram-se as tampas dos tubos, antes de os levar à ebulição;
12. Agita-se vigorosamente no *Vórtex* 5-10 segundos;
13. Centrifuga-se a 14000rpm durante 3 minutos;
14. Após extração, conservar as amostras a -20°C até posterior utilização.

Anexo II – Condições de amplificação do *Multiplex* PCR Master Mix (Qiagen)

	Tempo	Temperatura	Número Ciclos
Desnaturação Inicial	15 min	95°C	1
Desnaturação do DNA	30 seg	94°C	30
<i>Annealing dos primers</i>	90 seg	60°C	
Extensão da cadeia de DNA	90 seg	72°C	
Extensão final	15 min	72°C	1

Anexo III – Condições de amplificação do ABI Prism® SNaPshot™ *Multiplex* Kit (Applied Biosystems)

Número Ciclos	Tempo	Temperatura
25	10 seg	96°C
	5 seg	50°C
	30 seg	60°C

Anexo IV – Condições de amplificação dos kits utilizados para a análise HRM

SsoFast™ EvaGreen® Supermix (Bio-Rad)

	Tempo	Temperatura	Número Ciclos
Desnaturação Inicial	2 min	98°C	1
Desnaturação do DNA	5 seg	98°C	40
Annealing/ Extensão dos primers	15 seg	--- ¹	
Curva de Melting	10 seg/ step	65°C -95°C (inc. 0,2°C)	---

AccuMelt™ HRM SuperMix (Quanta Biosciences™)

	Tempo	Temperatura	Número Ciclos
Desnaturação Inicial	5 min	95°C	1
Desnaturação do DNA	5 seg	95°C	40
Annealing/ Extensão dos primers	30 seg	--- ¹	
Curva de Melting	10 seg/ step	65°C -95°C (inc. 0,2°C)	---

Type-it® HRM™ PCR (Qiagen)

	Tempo	Temperatura	Número Ciclos
Desnaturação Inicial	5 min	95°C	1
Desnaturação do DNA	10 seg	95°C	45
Annealing/ Extensão dos primers	30 seg	--- ¹	
Curva de Melting	10 seg/ step	65°C -95°C (inc. 0,2°C)	---

¹ A temperatura de *Annealing/ Extensão dos primers* depende do *primer* em estudo.

Anexo V – Proposta de artigo

COMPARING THE SNAPSHOT™ MINISEQUENCING AND HRM METHODOLOGIES IN THE CONTEXT OF FORENSIC GENETICS

Filipa Tavares¹ Luís Souto^{1,2} Helena Moreira¹ Fátima Pinheiro^{3,5,6} Francisco Corte-Real^{4,7}

¹Departamento de Biologia, Universidade de Aveiro

²CENCIFOR – Forensic Sciences Center

³Delagação do Norte, Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.

⁴Delagação do Centro, Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.

⁵ICBAS, Universidade of Porto

⁶Faculdade de Medicina, Universidade of Porto

⁷Faculdade de Medicina, Universidade de Coimbra

Abstract

In Forensic Genetics case work, there is a rising tendency to use mtDNA when limited quantities of DNA are available. The analysis of mtDNA in forensic samples is commonly carried out by direct sequencing of the hypervariable regions of the control region. However, the analysis of the control region has a limited power of discrimination. Genotyping SNPs in the coding region of mtDNA can provide additional information and increase the discrimination power of mtDNA typing. In recent years, new methodologies for the detection of SNPs have been developed. In this study, we compare two methodologies for the detection of SNPs in the control region of mtDNA: SNaPshot® and Real Time PCR, through High-Resolution Melting (HRM) analysis. Results showed that SNaPshot® is a more precise, robust and reliable methodology than HRM analysis. However, HRM analysis has the potential to function as a rapid and inexpensive pre-screening method for forensic samples prior to minisequencing.

Keywords: mtDNA; Coding region; SNPs; Real-Time PCR; Minisequencing; Forensic Genetics; East Timor population.

1. Introduction

Human mitochondrial DNA (mtDNA) is a useful marker for studying maternal biogeographic ancestry and is widely used in forensic casework and human population studies due to its specific features: 1) rapid mutation rate, 2) high copy number per cell, and 3) lack of recombination and maternal inheritance [1-3].

Most of the mtDNA sequence variations are located in the hypervariable segments I and II (HVS-I and HVS-II) of the control region [4]. For forensic identification, sequencing the HVS-I and HVS-II is the most common method for mtDNA analysis [5, 6]. However, the two routinely analyzed mtDNA hypervariable segments provide a limited power of discrimination in a forensic context. Since the sequencing of the entire mitochondrial molecule is not practical, the analysis of single nucleotide polymorphisms (SNPs) contained in the coding region of mtDNA seems to be a good strategy to overtake this drawback [7-9]. In recent years, interest in SNPs has increased and, simultaneously, numerous strategies have been developed to facilitate mtDNA SNP typing [10].

In the present study we compare two methodologies for the detection of SNPs in the control region of mtDNA: Minisequencing by SNaPshot® and Real Time PCR, through HRM (High-Resolution Melting) analysis. SNaPshot™ is one of the most common commercial technologies based on the minisequencing reaction and is a standardized methodology used for high-throughput SNP analysis. SNaPshot™ has been successfully applied in the genotyping of mtSNPs [8, 11]. HRM analysis is a post-PCR analysis used to identify genetic variations in a fragment of amplified DNA by comparing fluorescence in function of the temperature. Alleles produce distinct melting curves that can be compared with reference samples [12, 13]. The application of HRM analysis in the study of mtSNPs has been scarcely used, with exception to, for example [14].

2. Material and methods

2.1. DNA Samples

In the context of the genetic research of East Timor [15, 16], a total of 38 blood samples from the Southeast Asian collection, belonging to the Applied Genetics Laboratory at the Department of Biology (University of Aveiro), were analyzed. All analyzed samples were collected from unrelated donors who signed informed consent statements, and in agreement with local authorities – by Luís Souto.

Genomic DNA was extracted from whole blood using the Chelex®-100 modified method [17]. DNA concentration was measured using the Qubit® Fluorometer (Invitrogen™) and the samples were used with a final concentration of 5 ng.

2.2. SNP selection

The primers used in this study were selected to cover some of the haplogroups branches of the Asian mitochondrial tree, previously described in [9].

In order to facilitate samples analysis and increase peak resolution in HRM analysis, we chose the primers that could be analyzed in both methodologies (Minisequencing and HRM analysis). The length of the primer amplicons has a maximum of 109 bp and a minimum of 53 bp (Table 1). In this table, the primers sequence and annealing temperature are also described.

Positive and negative controls were used during the amplification and genotyping processes.

2.3. Minisequencing

2.3.1. PCR-multiplex amplification and purification

Multiplex PCR amplification was carried out in a total volume of 15 µl, containing 1x QIAGEN *Multiplex* PCR Master Mix (Qiagen), PCR primers in concentrations specified in Table 1 and 5 ng of DNA template.

Thermal cycling was carried out on Icyler (Bio-Rad), with the following conditions: 95 °C pre-incubation step for 15 min.; 30 cycles of 94 °C denaturation for 30 sec., annealing at 60 °C for 90 sec. and extension at 72 °C during 90 sec.; followed by 15 min. for final extension at 72 °C and 4 °C until removed from thermocycler. PCR products and negative controls were checked by agarose gel electrophoresis.

PCR products were purified to remove excess of primers and unincorporated dNTPs by adding 2 µl of ExoSAP-IT (USB Corporation) to each 5 µl PCR product. Reactions were incubated at 37 °C for 15 min. followed by 80 °C for 15 min. for enzyme deactivation.

Table 1: SNP coding region, primer sequences and final concentration for PCR amplification.

Nucleotide Position	Primer Sequence (5'→3')	Amplicon length (bp)	Final Concentration (μM)	Annealing Temperature (°C)
1119	GCTAAGACCCAAACTGGGATT GTTCTGGCGAGCAGTTTGT	84	0.5	56
2092	TGCCCACAGAACCCTCTAAA TGTCCAAAGAGCTGTTCTCT	69	0.5	60
4491	ATTAATCCCCTGGCCCAAC GATGAGTGTGCCTGCAAAGA	56	0.2	57
5417	CCTCAATCACACTACTCCCATA GGGGTGGGTTTTGTATGTTT	72	0.5	56
8281-8289del	TAGGGCCCGTATTACCTAT AAGAGGTGTTGGTTCTCTTAATCTTT	109	0.5	55
8414	TACTACCGTATGGCCACCA GGGAGGTAGGTGGTAGTTGTG	96	0.2	57
9824	TTTTGTAGCCACAGGCTTCC GAGGAAAGTTGAGCCAATAATGA	53	0.4	56
13759	GGAAGCCTATTTCGAGGATT GTTGTTTGAAGGGGGATG	64	0.4	57
12705	CCCAAACATTAATCAGTTCTTCAA TCTCAGCCGATGAACAGTTG	102	0.4	56

2.3.2. Minisequencing multiplex using SNaPshot and purification

The minisequencing reaction is performed using SNaPshot™ Kit (Applied Biosystems). The extension primers used in minisequencing have sizes between 25 and 67 bp; the length of the primers were modified with the addition of non-homologous tails poly(dC) or poly(dGACT) at the 5'-end (Table 2).

The minisequencing reaction was carried out using 3.5 μl of the SNaPshot™ Multiplex Ready Reaction Mix (Applied Biosystems), 1.5 μl purified PCR product, and 1.5 μl of extension primers mix (final concentrations on Table 2) in a 10 μl final volume. The reaction was performed on Icyler following the recommendation of the manufacture: 25 cycles of denaturation at 96 °C during 10 sec., annealing at 50 °C for 5 sec. and extension at 60 °C during 30 sec. To remove the unincorporated ddNTPs, the final product was treated with 1 μl of Shrimp alkaline phosphatase (SAP; Promega) at 37 °C for 60 min. followed by 80 °C for 15 min. for enzyme inactivation.

2.3.3. Electrophoresis detection

The minisequencing products (1 μl) were mixed with 10 μl of Hi-Di™ Formamide and 1 μl of GeneScan-120 LIZ (both Applied Biosystems). Samples were denatured at 95 °C for 5 min.; after denaturation, the samples are quickly cooled on ice for 5 min. Electrophoresis was run on an ABI PRISM® 310 Genetic Analyzer (Applied Biosystems) using Performance Optimum Polymer 6 (POP-6®). Resulting data were analyzed using GeneMapper® ID Software v3.2 (Applied Biosystems).

Table 2: SNP coding region, primer extension sequences and final concentration for minisequencing reaction.

Nucleotide Position	Extension primer ¹ (5'→3')	Lenght (bp)	Base substitution	Strand ²	Final Concentration (μM)
1119	(gact) ₈ GTTAGCCCTAAACCTCAACAGTAAA	58	T→C	L	0.40
2092	(gact) ₄ ACCCTCTAAATCCCCTTGTAATTTAA	43	C→T	L	0.50
4491	CCTGCAAAGATGGTAGAGTAGATGA	25	G→A	H	0.20
5417	(gact) ₉ CCATATCTAACACGTAAAAATAAAATGACA	67	G→A	L	0.40
8281-8289del	(gact) ₈ GACGTATTTACCTATAGCACCCCTCTA	61	C→G ³	L	0.30
8414	(gact) ₃ GACGCCACCATAATTACCCCATATA	37	C→T	L	0.20
9824	(gact) ₃ GCACAGGCTTCCACGGACT	31	T→C	L	0.20
13759	(gact) ₄ GATCATTACTAACAAATTTCCCC	40	G→A	L	0.15
12705	(c) ₁₇ AACATTAATCAGTTCTTCAAATATCTACTCAT	49	T→C	L	0.30

¹ Italic indicates the non-homologous tails.

² Strand refers to the target DNA chain for SNP genotyping

³ 9 bp deletion (8281-8289del) is interrogated as a C to G change.

2.4. Real Time PCR: HRM analysis

Reactions were carried out in 20 μl volume and contained 1 x SsoFast EvaGreen supermix (Bio-Rad), 0.3 μM of each forward and reverse primer (Table 1), and 5 ng of the DNA template.

Real Time PCR was carried on CFX96™ Real-Time PCR (Bio-Rad), with the following conditions: initial denaturation at 98 °C for 2 min., followed by 40 cycles at 98 °C for 5 sec. and 30 sec. at the primer's specific annealing temperature (Table 1). High resolution melting was carried out immediately after the PCR with the following conditions: 95 °C for 1 min. and 50 °C for 5 min., followed by an increase in temperature from 65 °C to 95 °C, with 0.2 °C increment, each step with 10 sec. hold.

The HRM data was analyzed using the *Precision Melt Analysis*™ Software. Raw fluorescence data was subjected to normalization and temperature shifting in order to remove background fluorescence, assist in visual interpretation and automatic grouping of similar melting curves.

3. Results and discussion

The results obtained for the two SNPs genotyping methodologies studied (SNaPshot® e HRM analysis) can be found in Table 3. When comparing the results obtained from the two methodologies, a 98% match was found. Discordance of results was marked (with an * in Table 3) for samples: 13 (SNP 1119); 28, 49 and 62 (SNP 8414) and 60 (SNP 13759).

Table 3: Comparison of results obtained in the two methodologies under study.

Sample ID	SNaPshot®										HRM analysis									Haplogroup ¹
	1119	2092	4491	5417	8281-8289del	8414	9824	13759	12705	1119	2092	4491	5417	8281-8289del	8414	9824	13759	12705		
	T	C	G	G	C	C	T	G	T	T	C	G	G	C	C	T	G	T		
1	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C	B4	
2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	
5	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	
8	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
11	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C	B4'5	
12	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
13	---	---	---	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---		
15	C	---	---	---	G	---	---	---	C	C	---	---	---	G	---	---	---	C	B4c	
16	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
18	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
19	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C		
20	---	---	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	F	
21	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
22	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
23	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C	B4	
25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	
26	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C	B4'5	
27	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C	B4	
28	---	---	---	---	---	---	---	---	C	---	---	---	---	---	*	---	---	C	R9c	
29	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	---	C	B4	
30	---	---	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	F	
31	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
45	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
46	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	
47	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	
48	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
49	---	---	---	---	---	---	---	---	C	---	---	---	---	---	*	---	---	C	R9c	
53	---	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---		
54	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
55	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
56	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
57	---	---	---	---	---	---	C	---	---	---	---	---	---	---	---	C	---	---	M7b-g	
58	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
60	---	---	---	---	G	---	---	---	C	---	---	---	---	G	---	---	*	C		
62	---	---	---	---	---	---	---	---	C	---	---	---	---	---	*	---	---	C	R9c	
72	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	
78	---	---	A	---	---	---	---	---	---	---	---	A	---	---	---	---	---	---	M9	
88	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	M*	

¹ Haplogroups were determined by the combination of markers studied in previous research as well as in this study.

SNP 8414 shows the largest inconsistencies between both methodologies. According to the SNaPshot® methodology, samples 28, 49 and 62 were recorded as wild type [C] while HRM analysis leads to a mutation (Figure 1).

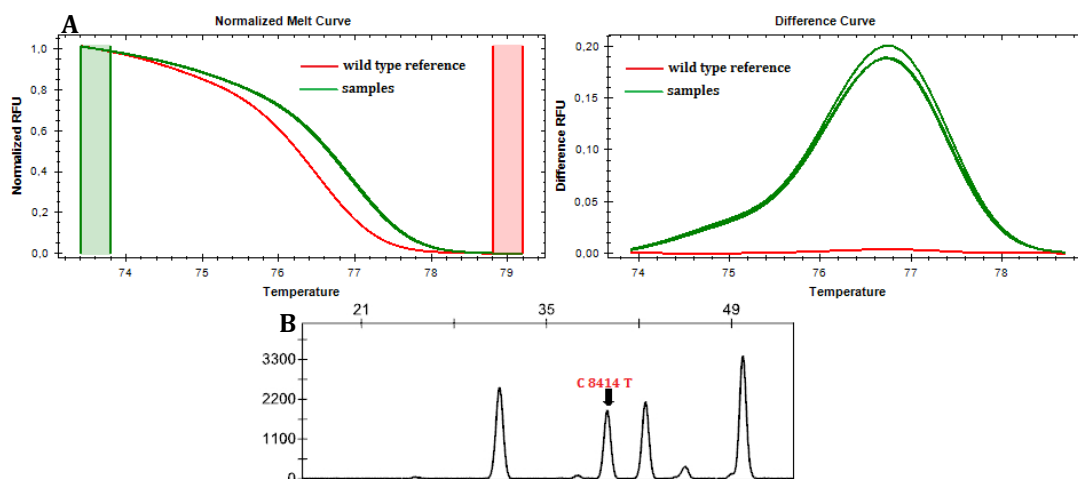


Figure 1 - Representative output files of samples 28, 49 and 62 (SNP 8414) in the evaluated methodologies. **A:** HRM results (normalized and difference curves), **B:** SNaPshot® readout after multiplex PCR.

Inconsistencies in SNP 8414 are due to the non-specificity of the primers in the HRM analysis: the primers used to detect SNP 8414 can amplify another polymorphism within the same amplicon. The samples 28, 49 and 62 were classified as subhaplogroup R9c based on other markers previously studied [16]. As a result, SNP 8414 cannot be present in the samples because it is specific to the haplogroup D4. Therefore, the difference in the melting curves of the samples may be the result of the existence of another polymorphism in the amplicons and not the presence of SNP 8414. Upon verifying the segment of the coding region of mtDNA generated by the primers used to detect SNP 8414 – through the comparison of primers with the reviewed reference sequence (rCRS) – it was verified that polymorphism 8440, present in the haplogroup R9c1a, was found in the amplicon. As a result, the polymorphism detected with HRM analysis could be 8440 (which classifies the samples in the subhaplogroup R9c1a) and not 8414.

The HRM analysis principle resides in the comparison of melting curves that have a particular shape according to the amplicon's composition of the nucleotide, without determining what polymorphism is present in a specific position. On the other hand, SNaPshot® is a method that uses extension primers that anneal to its target DNA immediately adjacent to the SNP in study. The fact that SNaPshot® uses specific primers for each polymorphism makes this methodology more specific in the detection of SNPs when compared to HRM analysis.

In order to improve the specificity of HRM analysis, it is important to use primers that amplify only one SNP per reaction. Occasionally, two or more SNPs may be located close to each other, making it difficult to design a primer to locate a single SNP. However, it is possible to amplify a region that contains multiple SNPs and carry out HRM analysis, as long as control samples for each SNP in study are included in the reaction.

Regarding the divergence of results in sample 13 (SNP 1119) and sample 60 (SNP 13759), this can also be explained by the non-specificity of the primers used in HRM analysis. Since it was not possible to determine the haplogroups of the samples, the polymorphism present in the amplicons was not identified.

4. Conclusions and forensic relevance

Analysis of mtDNA has proven to be a good alternative in forensic applications, when autosomal STR markers fail to give conclusive results due to the low amount of intact nuclear DNA. However, mtDNA analysis is frequently not being used due to high costs in sequencing the control region.

For this reason, rapid-screening procedures are of high interest, as they enable a pre-selection and reduction of samples that needed to be analyzed by sequencing. In addition, a SNP study can provide additional information and increase the discrimination power of the mtDNA typing.

SNaPshot® is a more precise, robust and reliable methodology, when analyzing multiple *loci* simultaneously. These advantages justify a greater use of this method in Forensic Genetics. In addition, it should be noted that the detection method uses a common technology as in current STR analysis (CE).

The disadvantages of HRM analysis, namely the fact that it does not offer information about the polymorphism present in the amplicon and is not easily applied in multiplex, makes this methodology more difficult to be applied in Forensic Genetics. However, as the amplification process with PCR is an essential part of the forensic sample analysis, HRM analysis can be integrated in laboratory routines as a method of sample pre-selection. Using HRM analysis, a high number of samples can be simultaneously analyzed in the identification of samples with interest for later SNaPshot® genotyping.

In summary, SNaPshot® is a more precise, robust and reliable methodology than HRM analysis. However, HRM analysis has the potential to function as a rapid and inexpensive pre-screening method for forensic samples reducing the number of samples that need SNaPshot® genotyping. The combination of these two methodologies can be a good alternative to overcome the low power of discrimination provided by the mtDNA control region sequencing.

Acknowledgements

We would like to thank the sample donors for their contribution to this study.

Our work was supported by the FCT Project: PTDC/CS-ANT/108558/2008 TIMORPOP GENÉTICA DE TIMOR-LESTE, Programa Operacional Temático, Fatores de Competitividade (COMPETE), Quadro Comunitário de Apoio III and FEDER.

References

1. Ballantyne, K.N., et al., *MtDNA SNP multiplexes for efficient inference of matrilineal genetic ancestry within Oceania*. Forensic science international. Genetics, 2011. **6**(4): p. 425-436.
2. Carracedo, A., *Forensic DNA typing protocols* 2005: Human Press Inc. 278.
3. Ingman, M., et al., *Mitochondrial genome variation and the origin of modern humans*. Nature, 2000. **408**(6813): p. 708-13.
4. Budowle, B., et al., *Mitochondrial DNA regions HVI and HVII population data*. Forensic Sci Int, 1999. **103**(1): p. 23-35.
5. Baasner, A., et al., *Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance*. Forensic science international, 1998. **98**(3): p. 169-178.
6. Barbosa, A.B., et al., *Mitochondrial DNA control region polymorphism in the population of Alagoas state, north-eastern Brazil*. J Forensic Sci, 2008. **53**(1): p. 142-6.
7. Köhnemann, S., et al., *A rapid mtDNA assay of 22 SNP in one multiplex reaction increases the power of forensic testing in European Caucasians*. Int J Legal Med, 2008. **122**: p. 517-523.
8. Quintáns, B., et al., *Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing*. Forensic Sci Int, 2004. **140**: p. 251-257.
9. Alvarez-Iglesias, V., et al., *Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups*. Forensic Sci Int Genet, 2007. **1**(1): p. 44-55.
10. Sobrino, B., M. Brión, and A. Carracedo, *SNPs in forensic genetics: a review on SNP typing methodologies*. Forensic science international, 2005: p. 181-194.
11. Brandstätter, A., T. Parsons, and W. Parson, *Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups*. Int J Legal Med, 2003: p. 291-298.
12. Reed, G.H. and C.T. Wittwer, *Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis*. Clin Chem, 2004. **50**(10): p. 1748-54.
13. Smith, B., C. LU, and J. Alvarado Bremer, *High-resolution melting analysis (HRMA): a highly sensitive inexpensive genotyping alternative for population studies*. Molecular Ecology Resources, 2010. **10**(1): p. 193-196.
14. Dobrowolski, S.F., et al., *Identifying sequence variants in the human mitochondrial genome using high-resolution melt (HRM) profiling*. Human Mutation, 2009. **30**(6): p. 891-898.

15. Souto, L., et al., *Y-STR haplotype diversity in distinct linguistic groups from East Timor*. American Journal of Human Biology, 2006. **18**(5): p. 691-701.
16. Souto, L., et al., *Mitochondrial DNA variability in populations from East Timor (Timor Leste)*. International Congress Series, 2006. **1288**: p. 115– 117.
17. Walsh, P.S., D.A. Metzger, and R. Higuchi, *Chelex-100 as a Medium for Simple Extraction of DNA for Pcr-Based Typing from Forensic Material*. Biotechniques, 1991. **10**(4): p. 506-513.